# Production of schwa by Japanese speakers of English: a crosslinguistic study of coarticulatory strategies

Yuko Kondo

A thesis submitted in fulfilment of the requirements
for the degree of Ph.D.
to the
University of Edinburgh
1995

EX BIBL. UNIV. EDINBURGH

*This thesis is dedicated to Jiro, Kumaji and Yumiko.*

# *Declaration*

*I declare that this thesis has been composed by myself and that the research reported therein has been conducted by myself alone.*

*Yuko Kondo*

*January 18th, 1995*

## Acknowledgement

I would like to thank all my supervisors, Alan Kemp, John Laver, Geoff Lindsey and Steve Isard, for their valuable comments and suggestions throughout the course. Geoff invented ingenuous sentences for the experiments. I owe thanks to Steve who has read countless earlier versions of the dissertation. I am particularly indebted to Mr. Kemp who has kindly volunteered to read through my manuscripts when Geoff left for his new career.

I owe special thanks to Gerry Docherty who suggested to me that schwa might be an interesting vowel to look at when I first plunged myself into the study of interlanguage phonology. Daniel Recasens has kindly sent me his data on Catalan schwa. Bob Ladd, Jim Scobbie and Harriet Magen have read parts of my dissertation and have given me important suggestions. I am also grateful to Ellen Bard and Jan McAllister for their advise on statistics.

I would like to thank all the computing and recording staff, Norman Dryden, Morag Brown, Irene Macleod, Stewart Smith, Cedric Macmartin and Leonard Scott. Without their patient assistance and support, this dissertation would not have been possible. I am also grateful to Prof. Motoko Yoshioka and the recording staff of International Christian University in Tokyo for letting me use the facilities. Thanks are also due to countless people who have participated in my experiments as subjects.

Michael Morony has helped me with the computer programs for the analyses. He has taken so much of his precious time for the programming. Interesting discussions I had with Mike and Sally Bates on vowels have always been a pleasure and encouragement. I am thankful to Jocelynne, Sandra, Eno, Nagita, Louise, Catriona, Mariko, Ethel, Irene, Mrs. Carter, Nucy, Keiko, and Li-yu for their morale support. In particular, Ethel and Irene have been so supportive both as staff members and friends. I would also like to express my special thanks to my office mates Dimitra and Etsuko. Their support made the last stage of my Ph.D. something special and fun.

I am also grateful to all the members of the Graded Direct Method Teachers' Association in Japan, in particular to late Miho Yoshizawa, Prof. Masukawa and Prof. Muro. Thanks are also due to Prof. Nakago and Prof. Murakami through whose guidance I became interested in phonetics.

For these long years, my parents and my husband's parents have waited for me with patience and understanding. They have taken such good care of Jiro and myself. I cannot express how much I have appreciated their warm support and encouragement.

Lastly but not the least, I must say a big "Thank You" to Jiro, Kumaji and Yumiko (and the rest of the cuddly fellows at home) for their great love and support.

# ABSTRACT


The rhythm of English is charaterized by the alternation of full and reduced vowels. Current studies on the reduced vowel /ə/ seem to suggest that schwa may be targetless in $F_2$ (Browman & Goldstein 1992c; Recasens 1986; Bergem 1993; Bergem 1994). This implies that there may be a contrast of targeted and targetless vowels in languages like English. In other words while the full stressed vowels of English are targeted and resistant to contextual effects (Fowler 1981; Magen 1984), the reduced vowels of English may be targetless and extremely variable as a function of contexts. This pattern of coarticulation seems to be determined at the higher level of the linguistic representation and surfaces as phonetic underspecification of schwa. On the other hand, there is no phonological vowel reduction in Japanese. Accent in Japanese shows little correlation with vowel quality. Presumably, all its vowels are targeted though large extent of V-to-V effects have been reported by a number of studies (Keating & Huffman 1984; Magen 1984). In other words, different accent types, stress accent and non-stress accent (Beckman 1986) observed in English and Japanese seem to constrain the coarticulatory patterns of the two languages in a global manner.

In the present study the coarticulatory pattern of the British English schwa was first observed by using VCəCV sequences with the contextual consonants of /p, t, k/ and the vowels /ɪ, æ, u/ embedded in natural sentences. The results of the study suggest that schwa may be targetless in $F_2$. I argue that this contrast in the extent of context dependent vowel variability is an important feature of stress-timing. In the second part, the coarticulatory pattern of the Japanese vowels was studied.

In the third part, variability of the English schwa produced by Japanese speakers of English was investigated. Two groups of Japanese speakers of English produced the VCəCV sequences described above. The two groups may be referred to as fluent and non-fluent non-native speakers of English. Similar VC_CV utterances were recorded in Japanese with the vowel /a/ as the middle vowel as it is the most likely candidate for a transfer from the L1 (the first language) vowel system.

The results showed that while the non-fluent group showed a pattern which may be characterized as a transfer from the Japanese vowel /a/ in $F_2$ and non-systematic variability in $F_1$, the fluent group showed a pattern similar to native speakers' production of schwa. They showed large and systematic variability in $F_2$ as a function of contexts. They seem to have acquired the phonetic under-specification of schwa.

Two of the fluent non-native speakers of English also showed stronger carry-over V-to-V effects in the labial context where most native speakers of Japanese would show stronger anticipatory effects in Japanese. The results seem to suggest that as L2 learners get to an advanced stage, they acquire the L2 phonological system which determines the coarticulatory pattern of the L2. The coarticulatory pattern of a language seems to be determined by a complex interplay of various constraints at different levels. Therefore, the above process seems to be far from an easy task, involving the restructuring of the system from L1 to L2.

# Contents

# Chapter 1

# Introduction

This dissertation addresses the question of how coarticulatory strategies differ from language to language and how such different and at times conflicting strategies may interact with one another when speakers produce second language (L2) sounds.

English and Japanese manifest an interesting contrast in coarticulatory strategies. The important claim of the present dissertation is that different coarticulatory patterns observed in English and Japanese may arise as a result of different accent systems employed by the two languages. In English there is a type of prosodic unit associated with the alternations between reduced and unreduced vowels, namely the stress foot (Abercrombie 1964; Selkirk 1980; Beckman 1986). Thus, accent plays an important role in organizing speech into prosodic units (see page 103). On the other hand, accent plays a minimal role in the prosodic organization of Japanese (Pierrehumbert & Beckman 1988; Kubozono 1987). This difference in the nature of accent may correlate with the different coarticulatory patterns observed in the two languages.

In English schwa occurs only in reduced syllables, that is, syllables with zero stress, and it is referred to as a reduced vowel. Current studies on schwa in English, Catalan and Dutch suggest that schwa may be phonetically unspecified (Browman & Goldstein 1992b; Browman & Goldstein 1992c; Recasens 1986; Bergem 1993; Bergem 1994; Koopmans-van Beinum 1994). In other words, schwa may lack a target value of its own and its phonetic value may be determined by the context alone. The concept is similar to the phonological analysis of French schwa by Anderson (1982) as an empty nucleus slot. That is, the empty nucleus is never filled in by any specification, but there is only a specified time interval between two segments in which the tongue continuously moves from one segment to another. Unstressed vowels in English have been observed to be more susceptible to contextual effects and thus are more variable as a function of contexts (Fowler 1981; Magen 1984). The phonetic underspecification of schwa makes a stronger claim in that it presupposes a contrast of targeted and targetless vowels. That is, while on the one hand there are stressed vowels that are prominent and resistant to coarticulatory effects, there may on the other hand be a completely targetless vowel that is a product of coarticulation. This seems to be a more economical strategy of speech production in that speakers will have to aim only at full vowels and leave the rest of vowels to contextual assimilation.

In most RP and conservative varieties of General American, there are two vowels that may occur in unstressed (zero stress) syllables and thus may be classed as reduced vowels. They are [ɪ] and [ə] in the words *happ***y** and *lett***er**. In non-final environments, however, there are some accents where the opposition between [ɪ] and [ə] has been lost while other accents retain the contrast as in RP *Lenin* ['lenɪn] vs. *Lennon* ['lenən]. Many accents are in a variable state regarding the contrast of [ɪ] and [ə] in non-final environments. An accent such as Australian has undergone the "Weak Vowel Merger" and has lost the contrast. Thus, the words *boxes* and *boxers* are pronounced identically [bɒksəz] and the words *rabbit* [ræbɪt] and *abbot* [æbət] rhyme in Australian. The vowel [ʊ] may also be classed as a reduced vowel for some accents and speakers; e.g., in the word *to* when it is unstressed (Wells 1982a; Wells 1982b; Kenyon 1946). In the present study, I limit myself to the study of schwa, the most frequent vowel in English, and will not go into a discussion of reduced vowels in general. However, Bates (forthcoming) has observed that the unstressed /ɪ/ is almost as variable as schwa due to contextual assimilation. Thus, reduced vowels in general may contrast with the full vowels in the extent of coarticulation.

In contrast to English, there is no report of accent dependent vowel variability in Japanese. Presumably, all its vowels are targeted. Nevertheless, Japanese vowels in general have been observed to coarticulate more strongly with contextual vowels than English full (stressed and weakly stressed) vowels (Magen 1984). This suggests that contextual variability is spread more or less evenly over all the vowels in Japanese to compensate for the lack of an unspecified vowel. In fact the studies reported here found that the extent of variability observed in Japanese vowels is intermediate in degree between that observed for the English full and reduced vowels.

The production of schwa by Japanese speakers of English is an interesting topic in that it implies the acquisition of a new coarticulatory strategy. The production of a native-like schwa entails the acquisition of the coarticulatory pattern of English. Japanese speakers of English will have to learn to produce targetless schwa and thus make a contrast of targeted and targetless vowels. This involves a shift from the L1 (the first language) coarticulatory pattern to the L2 coarticulatory pattern of the stress accent language.

The present study is structured in the following manner. In the first part I will argue that schwa is phonetically underspecified and that this feature of schwa plays an important role in the prosodic organization of English which defines the coarticulatory pattern of the language. In the second part, I explore the coarticulatory pattern of Japanese in detail. The important aim of this part is to empirically test whether there is any accent dependent vowel variability in Japanese. Finally in the last part, I will study the production of schwa by Japanese speakers of English. By observing the nature of interlanguage, i.e., a shift from the L1 to the L2 coarticulatory pattern, the characteristic features of both the L1 and L2 coarticulatory systems will become clearer.

The hypotheses to be tested in the present dissertation may be summarized as follows:

1. Schwa in English is phonetically unspecified, that is, its values are solely determined by its contexts.

2. The native speakers of English will manifest a sharp contrast in the extent of vowel variability between full and reduced vowels as a function of contexts.

3. The native speakers of Japanese will manifest no contrast in vowel variability between accented and unaccented vowels as a function of contexts.

4. The extent of vowel variability observed in Japanese will be intermediate in degree between that observed for English full and reduced vowels.

5. In shifting from the L1 to L2 coarticulatory system, speakers of L2 will show a 'merged' system (Flege & Hillenbrand 1984), that is, a system which is intermediate in value between the two systems. In the present case, Japanese speakers of English may manifest a contrast between the full and reduced vowels of English as a function of contexts. However, the difference in the magnitude of variability between the two types of vowels may be smaller than the difference observed for native speakers' production.

The basic assumptions underlying the present study are,

- The coarticulatory pattern of a language is largely determined and constrained by the phonology of the language.

- One of the important constraints that may affect the extent and the nature of the coarticulatory pattern of the language is its accent system.

- L2 learners may shift from the L1 coarticulatory pattern to a more L2-like pattern in the development of interlanguage. This implies the acquisition of the phonological representation of the L2. By acquiring the correct representation and learning to execute it, L2 learners may consequently acquire the phonetic detail of the L2.

## 1.1 An Overview

The present study consists of three main parts: the first part on English, the second part on Japanese and the third part on the interlanguage. The following is a brief overview of the structure of the present dissertation.

There are three chapters in Part 1. Chapter 2 presents a literature review on theories of underspecification and coarticulation as the background of the present study. It introduces some examples of the phonetic transparency of schwa and concludes by suggesting a number of constraints that may account for crosslinguistic phonetic differences observed in vowel variation.

Chapter 3 addresses the question of the phonetic underspecification of schwa. 'Targetless' schwa is still under debate and the issue has not been settled, though the Dutch data currently presented by Bergem (1994) lends a strong support to the theory. One of the important contributions of the present study is to provide further evidence in support of the theory. The present experiment is the first controlled experiment on British English schwa using real words in natural sentences. It also serves as reference data for the later experiment on interlanguage. It concludes that schwa may be partially targetless. That is, acoustically schwa may be targetless in $F_2$. The results support the view that vowel reduction may be contextual assimilation rather than centralization.

In Chapter 4, the extent of V-to-V effects on schwa and the full stressed vowel /æ/ of English are compared. The contrast of targeted (full) and targetless (reduced) vowels is illustrated. I argue that this contrast in the extent of coarticulation is an important feature of stress-timing. The existence of (an) unspecified vowel(s) seems to distinguish so called stress-timed languages from syllable-timed (e.g., Italian) and mora-timed (e.g., Japanese) languages. That schwa may retain its vocalic height or the status of syllable nucleus may also be an important feature of stress-timing.

There are four chapters in Part 2. Chapter 5 presents a literature review on the nature of the Japanese vowels. It gives a detailed account of the accent and the prosodic organization of Japanese following Pierrehumbert & Beckman (1988). An interesting point here is that in Japanese accent plays a minimal role in the prosodic organization of the language. This seems to correlate with the lack of accent dependent vowel variation in Japanese. A detailed account of vowel devoicing is also presented in this chapter as it is another important vowel weakening process widely observed in languages. It is also relevant to schwa as schwa is the target of the vowel devoicing process in English.

Chapters 6 through 8 present a series of experiments to explore the nature of vowel variation in Japanese. In Chapter 6, the effects of accent, pitch, phonemic vowel length and syllable position (initial and final in disyllabic nonsense words) on the vowel quality of Japanese are studied. As predicted from the nature of accent in Japanese, accent showed minimal (nearly nil) effects on the $F_2$ values of the Japanese vowels while accented vowels seem to be characterized by slightly higher $F_1$ values suggesting a more open articulation. Systematic effects of phonemic vowel length were observed.

In Chapter 7, the V-to-V effects on the Japanese vowel /e/ and /a/ are studied. An experiment is designed to show that accent does not affect the extent of V-to-V effects in Japanese. The extent and the nature of V-to-V coarticulation across the labial /b/ in Japanese are compared with those observed for the English /ə/ and /æ/ in Chapter 4. The relative strength of V-to-V effects on the English and Japanese vowels seems to be determined by the interplay of various constraints. The effects of secondary articulation (palatalization) on the extent of V-to-V coarticulation in Japanese are also observed.

In Chapter 8, the C-to-V and V-to-V effects on the Japanese vowel /a/ are studied. The experimental set-up replicates that of Chapter 3 designed for the English schwa. These experiments on schwa and the Japanese vowel /a/ serve as reference data for the experiment on interlanguage in Chapter 10. The choice of the vowel /a/ in this study is motivated as the vowel /a/ in Japanese is the most likely candidate for a transfer in the production of schwa by Japanese speakers of English.

There are three chapters in Part 3. Chapter 9 presents a literature review on theories in second language acquisition (SLA) studies. The review is not comprehensive in nature, but it covers the area that is most relevant to the data observed in the present study.

Chapter 10 presents an experiment on the production of schwa by Japanese speakers of English. Two groups of non-native speakers, fluent and non-fluent, produced schwa in VCəCV sequences with the consonantal contexts of /p, t, k/ and the vocalic contexts of /ɪ, æ, u/. These data were compared with the reference data produced by native British English speakers and the data on the Japanese vowel /a/. The two groups of non-native speakers of English showed different coarticulatory patterns. While non-fluent speakers showed a coarticulatory pattern very much like that of the L1 vowel /a/ in $F_2$ and non-systematic variability in $F_1$, fluent speakers showed a more L2-like coarticulatory pattern. They showed large and systematic context dependent vowel variability in $F_2$.

In Chapter 11, the nature of the coarticulatory pattern of the interlanguage is further explored by comparing V-to-V effects on the English /ə/ and /æ/

produced by Japanese speakers of English. The nature of V-to-V effects across the labial /b/ on the reduced and full vowels /ə/ and /æ/ of English produced by native and non-native speakers are compared together with the equivalent data on the Japanese vowel /a/. A number of observations regarding the relative strength in the directionality (L-to-R or R-to-L) of V-to-V effects in English and Japanese are summarized in this chapter.

Chapter 12 presents a general discussion and concludes the present study.

# Part I

# ENGLISH

# Chapter 2

# Theoretical Background

## 2.1 Introduction

While phonology deals with the grammar of the sound pattern of a language, phonetics has been considered to treat the physical properties of language sounds. Phonetics has thus been considered to be largely constrained by the physics of speech and universal in nature. However, a number of cross-linguistic and second language acquisition (SLA) studies have shown that languages manifest their typical ranges of values for some properties that are considered to be physically determined; for example, the height of velum for nasalization (Clumeck 1976) and voice onset times (VOT) for voiceless segments (Caramazza *et al.* 1973; Port & Mitleb 1981; Williams 1980; Flege 1981; Flege & Port 1981; Flege & Hammond 1982; Flege & Hillenbrand 1984). Gibbon *et al.* (1993) also showed that the extent of overlap between the gestures of /k/ and /l/ in the /kl/ sequence was different from language to language for six European languages, Catalan, English, French, German, Italian and Swedish. Though such fine differences are often not distinctive; i.e., they do not distinguish one meaning from another, and may be transcribed by the same broad phonemic category, nevertheless they contribute to the characteristic sound patterns of languages (Flege & Hillenbrand 1984; Hardcastle 1982; Disner 1983). As a result of these studies, a level of phonetic representation has been called for, and the mapping between the two levels of representations, phonetics and phonology has consequently been explored

(Journal of Phonetics vol.18 1990; Anderson 1993).

Underspecification and coarticulation are two areas in which the issue of the phonology-phonetics interface has been extensively studied (Pierrehumbert & Beckman 1988; Keating 1988; Cohn 1993). Coarticulatory phenomena range widely, from phonetics to phonology. At one end there are phonetic processes such as blurring and merging of adjacent segments in speech output. At the other end, there are assimilatory processes in phonology, e.g., feature spreading, vowel and consonant harmony. There are also casual or connected speech processes. Coarticulation, even at the phonetic end, seems to be largely planned (Whalen 1990), and is part of the grammar of a language.[1]

One might wonder if these are different phenomena, or different descriptions of the same phenomena. I consider that these are different realizations of the same phenomena, defined at different levels of the linguistic representation, phonological, syntactic, phonetic, pragmatic and sociolinguistic. The phenomena that underlie all these different realizations may be gestural overlap and zero gesture. I consider that a segment that is phonetically unspecified has a time slot with no gesture of its own. Its value is determined by the values of the contextual segments as described below. Most of other coarticulatory phenomena seem to be explained by the gestural overlap model advocated by Browman & Goldstein and their colleagues (see below for references).

Underspecification seems to play an essential role in coarticulatory phenomena. Often unspecified segments are considered to be transparent to coarticulatory effects. Keating (1988) describes an unspecified segment as simply interpolated through by the trajectory that moves from the preceding segment to the following segment. Such interpolation is considered to be phonetic in nature, as opposed to the phonological feature spreading where more categorical and discrete change in the nature of the affected segment is expected (see Figure 2.1 on page 17). In phonology, an unspecified segment generally receives a feature value

[1] Whalen (1990) asked speakers to read nonsense /aCV/ or /əbVCa/ strings (C = /p/ or /b/ and V = /i/ or /a/) before seeing the entire utterance. The segments known before the start of the articulation exhibited normal anticipatory coarticulatory influence, while those seen after the utterance onset did not. Carryover coarticulation was present for most cases in both conditions. The results seem to suggest that coarticulation is planned ahead of time, at least before the onset of a word.

in the course of a derivation, and thus is no longer unspecified in the output. In some cases, however, underspecification may persist onto the surface phonetic representation. This is the case of phonetic underspecification, and 'targetless' schwa seems to be such a case.

The following is a brief review of underspecification theories and coarticulation studies in order to capture the concept of phonetic underspecification of schwa and its implications on the crosslinguistic study of coarticulatory strategies aimed by the present thesis.

First, I will describe phonological underspecification as presented by Archangeli (1988) to introduce the essence of underspecification theories.[2] Then I proceed onto the description of phonetic underspecification as presented by Keating (1988). The key terms in both phonological and phonetic underspecification are *feature spreading* and *transparency*. A simple dichotomy of feature spreading as opposed to phonetic transparency presented by Keating (1988) will be introduced.

The key terms *feature spreading* and *transparency* play important roles in coarticulation studies as well. Some major speech production models that attempt to explain coarticulatory phenomena are introduced. There are two opposing views on coarticulation. They are feature spreading models represented by Henke (1966), Daniloff & Moll (1968), Benguerel & Cowan (1974) and others, and coproduction (gestural overlap) models advocated by Öhman (1966), Bell-Berti & Harris (1981), Fowler (1981), Browman & Goldstein (1986, 1990a, 1990b, 1992a) and their colleagues. Both models assume the transparency of segments to some features. The results obtained by Daniloff & Moll and Benguerel & Cowan, etc. seem to have been refuted by more controlled experiments conducted by Boyce *et al.* (1991) and Bell-Berti & Krakow (1991), and the feature spreading position seems to have been weakened. However, Boyce (1988) has shown that languages with different phonological rules, e.g., Turkish (with vowel harmony) and English (without vowel harmony) may exhibit different coarticulatory strategies. According to her, the coarticulatory pattern for Turkish was better explained as feature

[2]There are two major views in underspecification. Radical Underspecification and Contrastive Specification. Archangeli is a proponent of Radical Underspecification. See Archangeli & Pulleyblank (forthcoming) for more on this theoretical stance. For Contrastive Specification see Clements (1987) and Steriade (1987).

spreading while that of English fitted better with the coproduction model. It should be noted, however, that both transparency and feature spreading may be explained by coproduction model. In the case of feature spreading, the extent of gestural overlap may be greater and sustained with more muscular effort.

Jun & Beckman (1993, 1994) propose that vowel devoicing is a result of gestural overlap. According to their account, different degrees of vowel devoicing, the continuum of a fully voiced, partially voiced, and zero (completely deleted) vowel, may be correlated with different degrees of gestural overlap. For example, in the CVC sequence where the two C's are voiceless sometimes there are two separate gestures for the glottal opening, while at other times, these gestures come so close together that they become a single, long, sustained gesture (Figures 5.2 and 5.3 in Chapter 5). See Figures 3.14 and 8.5 for different degrees of vowel devoicing observed in English schwa and Japanese. Similarly, Boyce (1988) observed two separate gestures for lip rounding in the VCV sequence where the two [+round] V's are intervened by [−round] consonant in English. On the other hand, in Turkish, where vowels harmonize in [Roundness], she observed a unimodal activity of orbicularis oris, that is, a single long sustained gesture. The similarity of the two accounts is interesting in that feature spreading is accounted as a phonological process while vowel devoicing is accounted as a universal vowel weakening process. The key word here is a single sustained gesture. In the case of feature spreading, this type of gesture seems to suggest more effort. However, in the case of vowel devoicing, this single sustained gesture suggests less effort.

In any case Boyce's results suggest that coarticulation which is a physical property of spoken language may be phonologized in some languages and thus be manifested with different strategies. That is, coarticulatory phenomena themselves may range from what is described as phonetic transparency to phonological feature spreading. The nature of coarticulation therefore seems to vary from language to language. This variation in coarticulatory strategies needs to be accounted for at the level of both phonological and phonetic representations.

In the last section, some examples of phonetic transparency of schwa will be introduced. Its importance in the coarticulatory strategies of languages with phonological vowel reduction will be discussed. This leads to an argument that stress accent languages (with vowel reduction) and non-stress accent languages (without vowel reduction) may employ very different coarticulatory strategies. This difference in coarticulatory behaviour may be captured by observing schwa production by native (with vowel reduction) and non-native (without vowel reduction) speakers of English.

## 2.2 Underspecification

Theories of underspecification have recently come into prominence in both phonology and phonetics. The essence of underspecification theories is to supply predictable feature specifications by rule or default in the course of a derivation. At the level of phonological representation, feature values that are predictable remain unspecified in the feature matrix as illustrated below (Archangeli 1988). The two matrices are possible representations of a five-vowel system using the features [High], [Low], [Back] and [Voice]: (1)a is sparsely specified using the Radical Underspecification convention while (1)b is fully specified.

(1) a.

| | i | e | a | o | u |
|---|---|---|---|---|---|
| high | | − | | − | |
| low | | | + | | |
| back | | | | + | + |
| voice | | | | | |

b.

| | i | e | a | o | u |
|---|---|---|---|---|---|
| high | + | − | − | − | + |
| low | − | − | + | − | − |
| back | − | − | + | + | + |
| voice | + | + | + | + | + |

| | | |
|---|---|---|
| [+low] | → | [−high] |
| [+low] | → | [+back] |
| [ ] | → | [−low] |
| [ ] | → | [+high] |
| [ ] | → | [−back] |
| [ ] | → | [+voice] |

The selection of default rules is based on the concept of *markedness* in Universal Grammar though it is assumed that both universal and language-particular statements are needed to account for similarity and uniqueness observed among languages. For example, in the above example, devoiced vowels are universally marked, therefore all the vowels unspecified for [Voice] become [+voice] by default. Ideally, the specifications in the phonological representation of a language reflect phonological processes observed in the language. Yoruba shows asymmetry in regressive assimilation as shown below (Pulleyblank 1988). The vowel /i/ is the only exception to this process. This uniqueness of /i/ may be explained by representing /i/ as a featureless vowel from which no feature can spread onto the preceding vowel. Note that /i/ is completely unspecified in 1(a) above.

| | | | |
|---|---|---|---|
| (2) | owó Adé | owá Adé | 'Ade's money' |
| | owó ọmọ | owó ọmọ | 'child's money' |
| | owó ẹmu | owé ẹmu | 'wine money' |
| | arà ìlú | *arì ìlú | 'townsman' |

Underspecification is an essential concept to explain assimilation phenomena such as vowel harmony and consonant harmony. In vowel harmony features spread onto unspecified segments. For example, in Turkish vowel harmony, the possessive marker may surface as [u], [ɯ], [i] or [y] depending on the vowel in the stem to which it is attached as illustrated in (3). This phonological process is captured by considering this affix as unspecified for the features [Roundness] and [Backness]. The features [Roundness] and [Backness] spread from the vowel in the stem to the possessive marker.

| | | | |
|---|---|---|---|
| (3) | nominative | possessive | gloss |
| | top | top-u | *ball* |
| | kuʃ | kuʃ-u | *bird* |
| | | | |
| | kap | kab-ɯ | *lid* |
| | kɯz | kɯz-ɯ | *daughter* |
| | | | |
| | ɛt | ɛt-i | *meat* |
| | dil | dil-i | *language* |

| | | |
|---|---|---|
| dɛniz | dɛniz-i | *sea* |
| kœʃk | kœʃk-y | *villa* |
| tyrk | tyrk-y | *Turk* |

As an example of consonant harmony, a case of Sanskrit n-retroflexion is given in Archangeli (1988). In Sanskrit, a coronal nasal is retroflexed when it comes to the right of a retroflexed continuant, provided that no other coronal sounds intervene. Vowels and non-coronal consonants are both transparent to this rule. It is suggested that transparent segments are unspecified for the feature being spread. The noun forming affix *-ana* is added to verb roots in (4) (Whitney 1885; Whitney 1889).

| | | | | |
|---|---|---|---|---|
| (4) | rakṣ | protect | rákṣaṇa | protecṭion |
| | kṣubh | quake | kṣabhaṇa | |
| | vṛh | grow | vá̩rdhana | increase |

Transparency is a key concept in underspecification. In feature spreading, segments that intervene between the trigger and the target are transparent to the assimilatory process. Such segments are considered to be unspecified for the feature being spread. For example, consonants are generally transparent to the rules in vowel harmony.[3]

In theories of underspecification, it is generally assumed that phonetic material is given at the end of the derivation. However, an important issue in underspecification theories is a question of whether underspecification may persist onto the surface representation. It has been argued that segments that are phonologically unspecified for some features may remain unspecified at the level of phonetic representation (Pierrehumbert & Beckman 1988; Keating 1988).

Pierrehumbert & Beckman (1988) have shown that many of what used to be traditionally considered tone bearing units, such as syllable or mora, remain

[3]Vowels and consonants are also considered to be transparent to one another in phonetics. See page 23 on VCV coarticulation described by Öhman (1966).

unspecified even at the level of phonetic representation. Their key concepts are *target* and *interpolation*. The intonation contour is determined by a small number of targets that are aligned to certain prosodic units within a minimal contour unit. The $F_0$ values for the rest of the contour are determined by interpolating through these target specifications.

Phonetic underspecification seems to be a case of transparency. The phonetic transparency of an unspecified segment is clearest in the case of /h/ (Keating 1988). Being a glottal approximant, /h/ has no intrinsic oral gesture of its own. Spectrograms of intervocalic /h/ show that /h/ is simply interpolated through by the second formant trajectory that moves from the preceding vowel to the following vowel. This is in accordance with the phonological interpretation of /h/ as a segment with empty supralaryngeal articulation (Durand 1990).

A case of partial transparency has also been observed (Keating 1988). Observing spectrograms of VCV utterances where C is [s], she suggests that [s] may be specified for [Height] but unspecified for [Backness]. $F_2$ during [s] was observed to be moving from the value dominated by the first vowel to that dominated by the second vowel. On the other hand, $F_1$ was consistently low at the VC and CV transitions suggesting that [s] is specified as [+High], blocking V-to-V coarticulation in height.[4]

Phonetic transparency must be conceptually separated from *feature spreading* (Keating 1988). For example, /x/ in Russian illustrates a case of transparency when the $F_2$ trajectory in the sequence /ixa/ is observed. The $F_2$ trajectory during /x/ is dynamic and transitional. It moves gradually from the $F_2$ value of the preceding /i/ to that of the following /a/. On the other hand when /x/ occurs before /i/, it is fronted, and the fronted /x/ is a case of feature spreading in which the preceding segment takes on some property of the following segment. The fronted /x/ remains high in $F_2$ for most of its duration. For example, in the sequence /axi/, there is an abrupt transition from the preceding /a/ to /x/ and

[4]The effects of V-to-V coarticulation are generally observed less in $F_1$ than in $F_2$ (Kuehn & Moll 1972; Fowler 1981). As most consonants are characterized by narrow constrictions in the vocal tract and are generally specified as [+high], they may block V-to-V coarticulation in height.

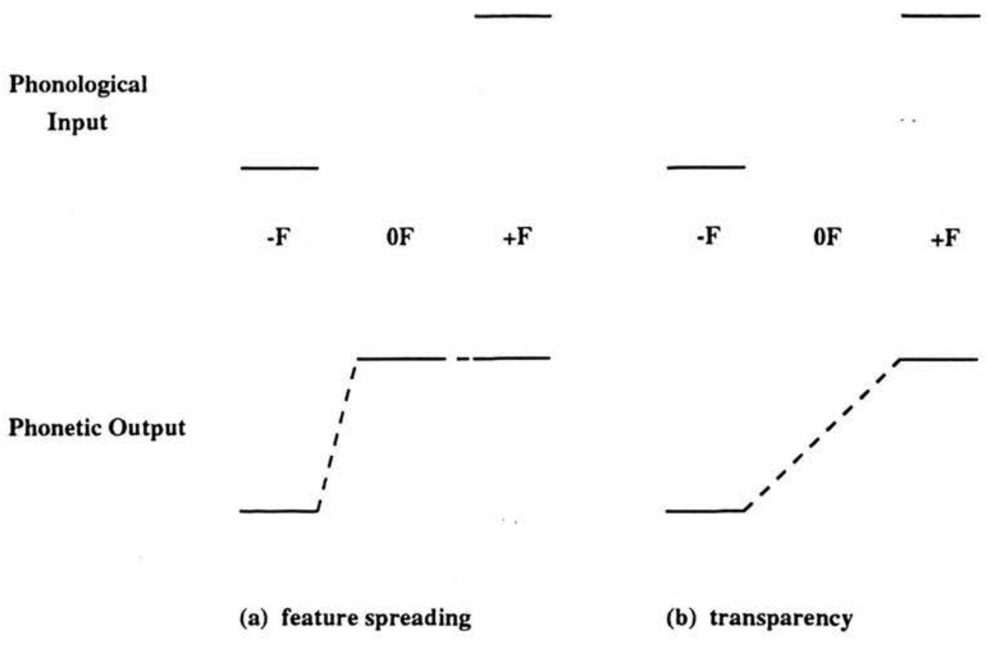



**Figure 2.1.** Schematic illustrations of feature spreading vs. transparency.

then $F_2$ stays more or less level till the end of the following /i/.[5] The difference in trajectory patterns for the case of transparency and feature spreading may be schematized as in Figure 2.1.[6]

However, the situation may be more complex than the simple dichotomy suggests. Boyce *et al.* (1991) and Bell-Berti & Krakow (1991) have shown that segments which lack specification for a certain feature, and thus are considered

[5]Asymmetry in VCV coarticulation was also observed across a velar consonant /g/ in Swedish by Öhman (1966). For the labial and alveolar consonants /b/ and /d/, such asymmetry was not observed; e.g., $F_2$ trajectories observed at the transitions of the vowels for /ybu/ and /uby/ were mirror images of one another. On the other hand, for /g/, the $F_2$ trajectory for /ygu/ seemed smooth and transitional, but for /ugy/, the continuity of the $F_2$ trajectory was not clear. This may be due to the palatalization effect of /y/ on /g/.

[6]Boyce *et al.* (1991), however, suggest an alternative explanation for the case of /axi/. They argue that the second vowel in /axi/ starts earlier, resulting in greater gestural overlap and the perception of the fronted /x/. Cohn (1993) argues that Nasal Spread in Sundanese is a case of feature spreading by saying that the nasal airflow traces are plateau like. However, her nasal airflow traces seem to be declining as a function of time. If the activation of the relevant muscle is unimodal, there would be a natural decline in the magnitude of the activity.


**Figure 2.2.** Schematic illustrations of the lip protrusion trajectories for $iC_nu$ and $iC_ni$ sequences observed in Boyce et al (1993).

to be transparent, may have characteristic articulatory positions for that feature. For example, vowels and oral consonants that are unspecified for [Nasality] are themselves associated with characteristic velum heights. Such characteristic positions may be often obscured or hidden in articulatory trajectories due to temporal constraints. Stretching a trajectory by introducing additional segments or reducing the speech rate, a slight rise in the trajectory due to a finer specification may surface. Figure 2.2 shows schematic illustrations of the lip protrusion trajectories for $iC_nu$ and $iC_ni$ sequences observed in Boyce *et al.* (1991). There is a single consonant between /i/ and /u/ in (a). The trajectory is smooth and the consonant between /i/ and /u/ seems transparent. However, in (b) where there is a multiple number of consonants, a local peak emerges as indicated by an arrow. In (c) a peak for C between two /i/'s may be observed where there is no [+round] feature to spread. A detailed phonetic observation suggests that segments which are considered to be transparent to a certain feature may still have some fine specifications for the feature concerned.

Boyce *et al.* (1991) also observed speaker specific articulatory positions. There may also be language or dialect particular characteristic positions for articulatory features. For example, vowels in the vicinity of nasal segments are nasalized to different degrees in different languages (Clumeck 1976). Though nasalization is caused by the physical constraint on the movement of the closing of the velum and considered to be a universal phenomenon, different languages have different degrees of velum opening. This points to the need of a language-specific fine grained phonetic grammar (Docherty 1989). Features are binary in phonology; either [+nasal] or [−nasal] and no intermediates. However, phonetic specification seems to be gradient. There is a whole range of velum heights from which a particular segment in a particular language may select its typical value. What is important is that these values are not only determined by physical constraints, but they are subtly controlled by the phonetic grammar of the language. Phonological processes are often described as categorical while phonetic processes are gradient. Browman & Goldstein (1992a) associate phonetic gradience with the nature of articulatory gestures while they associate categoricalness with the nature of speech perception. Clements (1992) expresses a similar view.

> Processes occurring during the act of talking will cause gradient changes that can ultimately be perceived as a categorically different gestural structure. (Browman & Goldstein 1992a:p171)
>
> Speech is produced in a gradient fashion, but perceived (and thus represented) categorically. (Clements 1992:p12)

## 2.3 Coarticulation

Phoneticians working on speech production have been intrigued by the gap between the abstract linguistic representation of discrete static invariant segments and the actual acoustic output which is continuously varying as a function of time. In the acoustic output segments are overlapping in time. They are continuous and dynamic in nature. There are no natural boundaries between adjacent segments but they influence each other, and their margins are blurred. The well-known and classic metaphor of this problem was presented by Hockett (1955:p210).

> Imagine a row of Easter eggs carried along a moving belt; the eggs are of various sizes and variously colored, but not boiled. At a certain point, the belt carries the row of eggs between the two rollers of a wringer, which quite effectively smash them and rub them more or less into each other. The flow of eggs before the wringer represents the series of impulses from the phoneme source. The mess that emerges from the wringer represents the output of the speech transmitter.

Studies dedicated to this problem may be generally referred to as coarticulation studies: studies of articulatory overlap. The term 'coarticulation' was coined by Menzerath & Lacerda (1933) to denote instances where two successive sounds were articulated together. The term is used interchangeably with the terms 'assimilation' and 'feature spreading', but covers a wider range of phenomena (Ohala 1993), for as far as acoustic outputs are concerned, it is harder to find a case where segments are not coarticulated. A number of speech production models have been proposed to account for coarticulatory phenomena, such as Henke's feature-based model (Henke 1966), hierarchical models such as syllable-based model (Kozhevnikov & Chistovich 1965), and coproduction models (cited on page 11) among others. Underspecification has been a central issue in these models. Feature spreading and transparency are again crucial concepts in coarticulation studies.

The idea of a feature-based model originated in Henke's computer model of the English stop + vowel articulation. This model is basically a feature spreading model. It is known as a 'look-ahead' model for anticipatory coarticulation. The idea is that, in speech production, the look-ahead mechanism looks for a segment which is specified for a certain feature, e.g., [+round], and all the preceding segments that are unspecified for this feature take on the feature [+round] until the spreading of the feature is blocked by a segment which is specified otherwise; in this case [−round]. The idea presented here is remarkably similar to the case of Sanskrit n-retroflexion presented above.

Kozhevnikov and Chistovich's model also implies feature spreading. According to their model, the basic unit of speech production in Russian is a syllable: a vowel and any number of preceding consonants. The plan for an entire syllable is executed at the onset of the syllable. In this sense, the model is a coproduction model; that is, all the segments within a syllable are coproduced. The model also suggests hierarchical structures in speech production by setting the domain for feature spreading. That is, features on a vowel such as [+round] or [+nasal] may spread onto the preceding consonants within a syllable but not beyond. Within a syllable, as long as there is no contradictory articulatory specifications, features spread onto the preceding segments.

Coarticulation is described as gestural overlap in articulatory phonology as proposed by Browman & Goldstein (1986). There is a strong parallel between articulatory phonology and feature geometry (Clements 1985). The modularity of features are captured in these phonological representations that may be considered as the extension of the principles of autosegmental phonology (Goldsmith 1979). The essence of autosegmental phonology is that features are autonomous and independent of one another. What is common in these representations is that features are "articulator-bound" (Ladefoged & Halle 1988) and that they should be classed by the articulator that executes them: larynx, soft palate, lips, tongue blade, tongue body, tongue root (Browman & Goldstein 1990b). As these features or gestures are on different tiers, they may be independent of one another, and they may overlap in time. Transparency is guaranteed by the modularity of features. Most assimilatory processes may be accounted for by gestural overlap according to this model (Browman & Goldstein 1990b; Zsiga 1993; Jun & Beckman 1993; De Jong *et al.* 1993). The same-tier overlap, however, is still possible. Munhall & Löfqvist (1992) have shown the blending of glottal gestures for /st/ of *Kiss Ted* at fast tempi. Browman & Goldstein (1990b) also illustrate cases of blending such as the assimilatory process in a sequence /nθ/ of *tenth* as [n̪θ].

A number of researchers have looked into the nature of feature spreading (Daniloff & Moll 1968; Moll & Daniloff 1970; McClean 1973; Benguerel & Cowan 1974; Bell-Berti & Harris 1979; Bell-Berti & Harris 1981). Lip rounding was observed to span as far as six segments from the trigger vowel in a nonsense sequence such as /istrstry/ in French (Benguerel & Cowan 1974). Moll & Daniloff (1970) observed that nasality spreads across a word boundary in a sequence such as *free Ontario.* McClean (1973) observed nasal coarticulation in English across different syntactic boundaries. He found that low-level word boundaries did not usually constrain coarticulation though high-level boundaries such as those between phrases, clauses or sentences did.

However, Bell-Berti & Harris (1981) have concluded that lip rounding begins at a constant time before the trigger vowel. The EMG activity from the orbicularis oris muscle, responsible for lip rounding and protrusion, was observed to begin at a constant time before the beginning of the acoustic period of the [+round] vowel. The model presented by Bell-Berti & Harris (1981) is basically a coproduction model. The lip rounding gesture of a rounded vowel is coproduced in time with other articulatory gestures of the preceding segments.

The data presented by Boyce (1990) (discussed briefly on page 17) favour the coproduction model. Comparing lip protrusion in the sequences /i$C_n$i/ and /i$C_n$u/, using an optoelectronic tracking system (modified Selspot), rounding was observed on the consonants before the [−round] vowel /i/ as well as before the [+round] vowel /u/. It was concluded that segments unspecified for a certain feature have characteristic positions for that feature. Similar observation was made for the spread of [Nasality] (Bell-Berti & Krakow 1991). Thus, the lip rounding observed by Benguerel & Cowan (1974) on the consonants /strstr/ before a rounded vowel /y/ (discussed above) may not be due to the spreading of the feature, but may be due to the characteristic lip protrusion associated with the consonants that are considered to be unspecified for the feature (Figure 2.2).

It is plausible that phonetic assimilation is time-locked to the trigger segment while the domain of phonological feature spreading may be rule-governed. For example, Ghazeli (1977) has observed that a word is the domain for tongue-backing assimilation in Arabic. Jun's (1993) data on Korean vowel devoicing suggest that an accentual phrase may be the domain of coarticulation in Korean.[7] Fowler (1981) suggests that in English speakers may organize the intervals between stressed syllables within a phrase as a basic unit of speech. This interval may possibly be the domain of coarticulation in English. Boyce (1988) compared lip rounding coarticulation between languages with and without phonological feature spreading (see also page 11). In Turkish, there is a vowel harmony rule for the spreading of the feature [Roundness] while in English there is no such rule. She showed that the coarticulatory pattern characterized by the articulatory trajectories and EMG activities for Turkish fitted better with the feature spreading

[7]There are three way contrasts of voicing in Korean; aspirated, lenis and fortis. An aspirated stop has a large glottal opening gesture, the peak of which is aligned around the oral release. A fortis stop has a much smaller opening that ends in a tightly closed glottis before the oral release. A lenis stop has a large glottal opening like an aspirated stop in initial positions, but in medial positions there is often no apparent opening between the vocal folds and as a result voicing is observed. Jun reports that the voicing assimilation of a lenis stop is observed within an accentual phrase, i.e., either when it occurs word-medially or word-initially within an accentual phrase. A lenis stop is aspirated only when it is at the beginning of an accentual phrase. An accentual phrase is marked by an initial peak accent (in Chonnam Korean) or by a final rising accent (in Seoul).

model whereas the pattern for English could be better described by the coproduction model. In the lip protrusion trajectories obtained by an optoelectronic tracking system, she observed a trough between two rounded vowels for English speakers suggesting that lip rounding is time locked to the trigger vowel. She also observed bimodal patterns in the EMG activity of the orbicularis oris muscles for English speakers. On the other hand, for Turkish speakers, a more plateau like trajectory without a trough was observed, and their patterns of EMG activity was unimodal rather than bimodal, suggesting that feature spreading is a better account for lip rounding observed in Turkish. It is suggested that different languages may employ different coarticulatory strategies motivated by the phonology of the language. In Turkish, the phonological rule of vowel harmony seems to favour a feature spreading rather than a coproduction strategy. It should be noted, however, as mentioned earlier that both types of coarticulatory patterns may be interpreted as a case of gestural overlap. That is, in the case of feature spreading, the gesture for that particular feature is sustained with more effort. In any case, Boyce's data suggest an interesting shift from what is a universally and physically constrained phonetic process to a somewhat arbitrary and rule-governed phonological process.

Öhman (1966), in the classic study of VCV coarticulation, observed the effects of the first vowel at the onset of the second vowel, and the effects of the second vowel at the offset of the first vowel. The $F_2$ trajectory looked as though there was a single continuous movement from $V_1$ to $V_2$ across the middle consonant. He concluded that the consonant articulation is superimposed on the basic diphthongal articulation of the vowels.[8] That is, consonants and vowels are coproduced. This suggests the transparency of consonants to vocalic gestures.

However, the transparency of consonants to vowels is not absolute. Some consonants seem to be more transparent to the V-to-V articulation than others. For example, [h] seems to be completely transparent to the V-to-V articulation (Keating 1988). On the other hand, palatal consonants seem to block the V-to-V coarticulation. Coarticulatory resistance of different consonants has been

[8]Similarly, in CVC utterances, the acoustic effects of the first consonant were observed at the VC and the effects of the second consonant were observed at the CV transition (Broad & Fertig 1970).

studied by a number of researchers (Bladon & Al-Bamerni 1976; Recasens 1984; Farnetani 1990; Recasens 1991; Farnetani & Recasens 1993). Recasens (1984) reports that consonants with different degrees of tongue dorsum activity, [j] > [ɲ] > [ʎ] > [n], exhibit different degrees of V-to-C coarticulation, [n] > [ʎ] > [ɲ] > [j]. The relationship is inversely correlated. That is, the more involved the tongue dorsum is in the articulation of a consonant, the more resistant it is to the coarticulatory effects from the adjacent vowels, and thus it reduces or blocks the V-to-V coarticulation. Öhman (1966) observed less anticipatory V-to-V coarticulation in $F_2$ for Russian compared to English or Swedish. The presence of contrastive secondary articulation, a contrast of palatalized vs. non-palatalized consonants, seems to block V-to-V coarticulation in Russian. Keating (1985) argues that consonants in Russian are specified for vowel features and thus are opaque to V-to-V gestures.[9]

Keating (1990) has incorporated a theory of underspecification into a model of coarticulation. According to the model each feature value of a segment such as [Nasality], for the opening of the velum, or [Height], for jaw position, is associated with a range of possible articulatory or acoustic values, i.e., a minimum and maximum value within which all the observed values must fall. This range of values is called a "window". A window reflects the sensitivity of a segment to its contexts.

> For some segments this window is very narrow, reflecting little contextual variation; for others it is very wide, reflecting extreme contextual variation. Window width thus gives a metric variability. There is no other "target" associated with a segment; the target is no more than this entire contextual range (Keating 1990:p455).

[9] Japanese has five phonemic vowels with a contrast of long and short vowels. It is a non-stress accent language. It has a phonemic contrast of palatalized vs. non-palatalized sounds; e.g., /ka/ vs. /k$^j$a/, /ku/ vs. /k$^j$u/, and /ko/ vs. /k$^j$o/ before [+back] vowels for the segments /p, b, t, d, k, g, s, z, m, n, r, h/. However, unlike Russian, strong V-to-V coarticulation is observed (Magen 1984; the present study). This suggests that the presence of contrastive secondary articulation is not sufficient to explain the smaller magnitude of V-to-V coarticulation observed in Russian and other Slavic languages (Choi & Keating 1991). As suggested in Section 2.4 there may be a more complex interaction of various constraints in determining the extent of coarticulation observed in a language.

Articulatory or acoustic trajectories move through windows. This model is an extension of the targets and interpolation model. Under this model, however, by introducing the concept of window, the nature of phonetic underspecification is considered as gradient, rather than being a discrete contrast of "specified" and "unspecified". A segment is *less* specified when its window is wide, while a segment is *more* specified when its window is narrow, and all intermediate degrees are possible. Thus, different degrees of transparency observed in consonants against V-to-V coarticulation may be accounted for by assigning different window sizes to different consonants. For example, /h/ may have a very wide window, whereas /j/ has a narrow window. In this sense, windows allow greater degree of freedom for target interpolation.

In discussing the nasalization of a vowel between two nasal consonants, i.e., NVN, Keating (1990) suggests that the vowel window for the velum height excludes a straight line interpolation between the two nasals though vowels in English are phonologically unspecified for [Nasality]. A slight rising of the velum was observed by Kent *et al.* (1974) and Vassière (1983) during a vowel between two nasal consonants. Stevens (1990) suggests that the place cue of the nasal consonant is perceived better when the extent of the nasal coupling is reduced around the consonantal release or implosion. This favours a higher position of the velum during a vowel. Characteristic positions for the velum height observed for the segments that are unspecified for [Nasality] (Boyce *et al.* 1991; Bell-Berti & Krakow 1991) may be due to such physical motivations. Such characteristic positions may be defined by the window size and position. Thus, the gradience of phonetic specifications at the level of the surface representation may be explicitly expressed by the window model.[10]

## 2.4 Transparency of schwa

Schwa is a good candidate for underspecification. Figure 2.3 shows two phonetic manifestations of one underlying phonological entity /ə/ in the utterance

[10] A computer model with window like targets by F. Guenther of Boston University is in press in Biological Cybernetics.

*I'm dreaming* **of a** *White Christmas.* The difference is dramatic. Figure 2.4 is a spectrogram of the utterance *orb* **a***borted* where the first intervocalic /b/ is lenited and the schwa is almost completely interpolated through by the $F_1$ and $F_2$ trajectories of the neighbouring vowel /ɔ/. Huffman (1986) reports that the effect of a first vowel on a third was observed in $F_2$ across two /l/'s and /ə/ in VCəCV utterances. This suggests the transparency of schwa in V́-to-V́ articulations. Flege (1988) observed the tongue movements in the productions of *a bib again* and *a bob again* using glossometry. In the production of *a bib again*, the tongue moved from /ɪ/ slightly downward to the poststressed schwas and up again for the following /g/ of *again*. However, following /ɑ/, though one subject stopped for schwa, another subject moved the tongue a great distance up towards a configuration associated with the /g/ without stopping for the schwa. In either case schwa could be heard on the audiotape. Such examples of apparent underspecification of schwa are frequently observed. This observation of schwa as an unspecified segment is in accordance with the phonological analysis of schwa as an unspecified vowel [+syll, -cons] with zero supralaryngeal articulation (Durand 1990).

Targetless schwa would be a product of contextual assimilation. It would be transparent to its surrounding sounds, and its value would be determined by context alone. Then a range of variation for schwa is expected to cover the whole range of its possible contexts. Recasens (1991) has shown that schwa in Catalan is more variable than any other vowel as a function of different consonantal contexts in his articulatory (EPG) and acoustic ($F_2$) studies. Koopmans-van Beinum (1994) has also shown that in Dutch, the standard deviation of the schwa $F_2$ is greater than that of any other vowel. On the other hand, when the standard deviations of $F_1$ for Dutch vowels are compared, schwa takes an average position between all the vowels. In general the standard deviations increased with the greater degree of openness of the mouth. This implies that schwa may be targetless in $F_2$, but targeted in $F_1$.

If schwa is completely targetless, its articulatory and acoustic trajectory would move continuously from the preceding segment to the following segment without going through any target of its own. Magen (1989) reports that in trisyllabic sequences with a medial schwa, [bVbəbVb], $F_2$ moves roughly continuously from

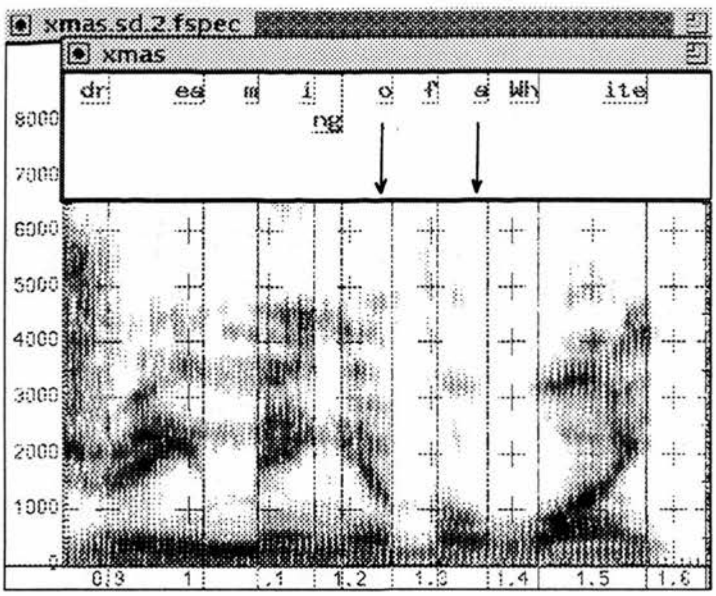



**Figure 2.3.** Two phonetic manifestations of the phoneme /ə/ in the utterance *I'm dreaming* **of a** *White Christmas* by an adult male speaker (GL) with standard Southern British English accent.

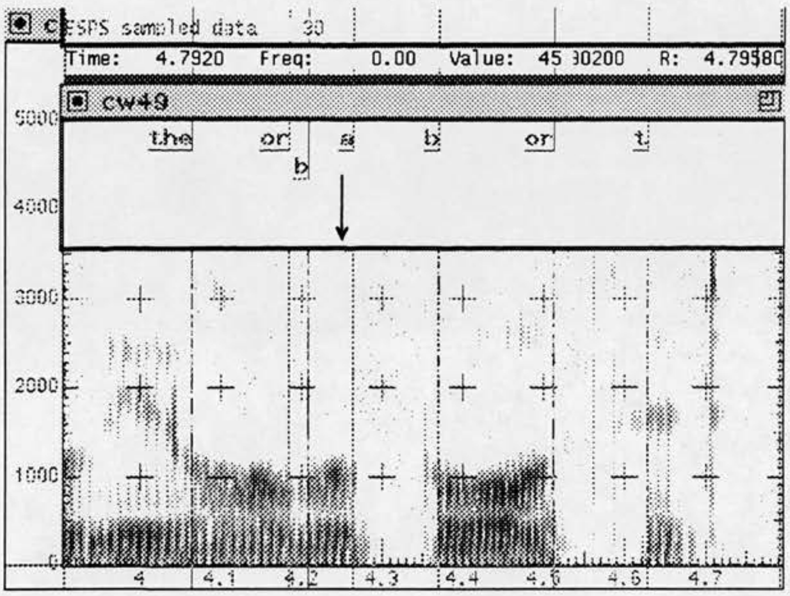


**Figure 2.4.** The spectrogram for the sequence *orb aborted* by an adult male speaker (CW) with standard Southern British English accent.

the value dominated by the preceding vowel to that dominated by the following vowel across intervocalic consonants and through the schwa.

The transparency of schwa must be defined at the higher level of the linguistic representation and persists onto the surface level. This transparency of schwa has an important implication for coarticulatory strategies of languages with and without stress accent (Beckman 1986). So called stress-timed languages are characterized by phonological vowel reduction. If schwa is targetless, the contrast of full and reduced vowels may be described as a contrast of targeted and targetless vowels. Speakers need to aim only at targeted, full vowels and leave the rest of the vowels to contextual assimilation. On the other hand, in a non-stress accent language like Japanese with no such contrast, all its vowels are presumably targeted. Nevertheless, a large extent of vowel variation is observed in Japanese (Keating & Huffman 1984; Magen 1984). This suggests that contextual variability is spread more or less evenly over all the vowels in Japanese to compensate for the lack of an unspecified vowel. This might imply an essential difference in coarticulatory strategies between stress and non-stress accent languages. Vowel variation may thus be constrained by the accent systems of languages.

Manuel & Krakow (1984) propose that the extent of vowel variation observed in a language is constrained by the number and distribution of vowels in the language. That is, vowels are less variable when they are more crowded in the acoustic vowel space in order to avoid overlap and maintain their identities. They have shown that vowels in Swahili, a five-vowel Bantu language, are more variable than vowels in English, a language relatively rich in vowels. Manuel (1990) has also observed that Sotho, another Bantu language with seven vowels shows less vowel variation than Nbedele or Shona with a five-vowel system, also of the same Southern Bantu family. According to their hypothesis, the constraint on vowel variation is physically (or perceptually) motivated. The two types of constraints, the accent system and the number and distribution of vowels in a language, are not mutually exclusive. It is plausible that different constraints are at work in determining the extent of vowel variation in a language. Other possible constraints are the existence of secondary articulation such as palatalization in Slavic languages (Choi & Keating 1991) and the existence of phonological feature spreading rules such as vowel harmony in Turkish.

Further, these constraints may be ranked as proposed in Optimality Theory (Prince & Smolensky 1993). The basic idea explored in Optimality Theory is that Universal Grammar consists largely of a set of constraints on representational well-formedness from which individual grammars are constructed. The constraints operating in a particular language are, however, highly conflicting and make contrary claims about the well-formedness of most representations. The grammar therefore consists of constraints together with the means of resolving conflicts between various constraints. The heart of the proposal of this theory is "a means for precisely determining which analysis of an input best satisfies (or least violates) a set of conflicting well-formedness conditions." The grammar rates all these analyses and places the one that best satisfies the whole set of constraints at the top of the list. This is the *optimal* analysis. "Optimality Theory relies on a conceptually simple but surprisingly rich notion of constraint interaction whereby the satisfaction of one constraint can be designated to take absolute priority over the satisfaction of another. The means that a grammar uses to resolve conflicts is to rank constraints in a strict dominance hierarchy." (Prince

& Smolensky 1993:Chap. 1). The Optimality Theory may give a phonological account of gradient phenomena in phonetics.

The acquisition of schwa by Japanese speakers of English is interesting in that it involves a shift from one coarticulatory pattern to another. As the coarticulatory pattern of a language may be determined by a complex interplay of various constraints, which are different from language to language, such a shift may be far from a simple process. The present study attempts to understand the nature of this shift, focusing on the difference in the accent system of L1 and L2.

# Chapter 3

# Phonetic Underspecification of Schwa

## 3.1 What is schwa?

Vowel reduction has traditionally been considered as centralization. A tube modelling a centralized configuration of an average sized male speaker resonates at 500 Hz, 1500 Hz, and 2500 Hz (Fant 1960). That is, the formants are evenly spaced. Lehiste (1970) suggested that in both British and American varieties of English, most vowels approach a vowel of this central quality, namely schwa, when they are in weakly stressed syllables. Schwa is described as a central vowel by Kenyon (1946), Gimson (1980) and Crystal (1985).

On the other hand, unstressed vowels have been observed to coarticulate more strongly with the surrounding segments than stressed vowels (Magen 1984; Fowler 1981). Schwa, in particular, is observed to be most susceptible to coarticulatory effects (Magen 1989; Recasens 1986; Recasens 1991; Bergem 1993; Bergem 1994; Koopmans-van Beinum 1994; Bates forthcoming). Kenyon (1946) also notes that the quality of schwa is much affected by surrounding sounds. It is also described as an 'indeterminate vowel' by Clark & Yallop (1990).

From these observations, a question arises: is vowel reduction centralization or contextual assimilation? Both centralization and contextual assimilation imply target undershoot. In either case, vowels do not reach their target value and result in a value more like their surrounding segments (assimilated) or having a value less extreme or peripheral (centralized).[1]

Nord (1974) concluded that vowel reduction is a result of increased contextual assimilation and that schwa is central in quality only when it is assimilated to the neutral vocal tract configuration of a rest position; that is, when it is adjacent to a pause.

Another advocate of the contextual assimilation hypothesis is Bergem (1993, 1994). He argues that 'centralization' could be a result of averaging different contextual effects on schwa.[2] He suggests that Dutch schwa may be interpreted as a vowel without articulatory *target* and that it is completely assimilated with its phonemic context.

On the other hand, there seems to be a case of centralization per se as a result of vowel reduction. Stevens & House (1963) have shown that vowels in consonantal contexts have more central $F_2$ values compared to vowels in isolation. However, Stevens & House (1963) did not use consonants such as a palatal /j/ or a velarized /l/ which might have perturbed the vowel formant values to more peripheral values.

Segments that are shorter in duration also appear to centralize. Keating & Huffman (1984) observed that phonemically short /o/'s in Japanese were more central in the acoustic vowel space than their long counterparts, /oo/'s. However, these observations may also be interpreted as a result of increased contextual assimilation. Lindblom (1963:Fig. 2, p1775) has shown that as a vowel gets shorter in duration, its formant frequency values at the steady-state approach those of the contextual consonants /b, d, g/ observed at the initial transition.

[1] Lindblom (1990) has suggested that there is a continuum in the style of speech ranging from *hyper-* (careful) to *hypo-* articulated (sloppy) speech. Johnson *et al.* (1993) showed that phonetic targets of vowels as they exist in our mental inventory of phonemes are hyperarticulated. Listeners were asked to adjust the input parameters of a speech synthesizer until the vowels it produced sounded like the vowels in a set of example words. The vowels that listeners chose were more peripheral, that is, high vowels were higher, low vowels were lower, front vowels were farther front, etc., than the vowels which they were asked to synthesize. It would be interesting to test how listeners would synthesize schwa.

[2] The mean of all the peripheral values obviously lies at the centre of these values. In this sense the results obtained by Browman & Goldstein (1992c) is suggestive. The mean pellet positions for schwa in their X-ray microbeam study lay almost on top of the grand means for the pellet positions across all the full vowels. See below for more discussion.

This trend was observed for $F_2$ and $F_3$, and was particularly pronounced for $F_3$ (Figure 3.1).

The present study addresses the question of whether vowel reduction is centralization or contextual assimilation by testing the following hypothesis.

> Schwa is phonetically unspecified: that is, it has no inherent value of its own, but its acoustic values are determined by its contexts.

Firstly, if schwa is a product of contextual assimilation, the range of its values will cover the whole range of the values of its possible contexts. On the other hand, if schwa is a vowel of central quality with a target, then the range of its values will be limited and will cluster around the central value. Secondly, if schwa is targetless with no value of its own, the schwa segment will simply be interpolated through by the formant trajectories that move from the preceding segment to the following segment as illustrated below (Figure 3.2). In symmetric contexts, the trajectories across schwa will be level.

In an X-ray microbeam study of schwa articulation and a study of computer-simulated schwa, Browman & Goldstein (1992c) conclude that schwa in the nonsense utterances [p$V_1$pəp$V_2$pə] is targeted. Schwa was observed to be more or less interpolated through when $V_1$ and $V_2$ were unlike one another as in the utterance [pipəpapə], but a warp in the trajectory was observed when schwa was between two like vowels as in [pipəpipə]. The results of multiple regression analyses showed that the schwa trajectory connecting $V_1$ and $V_2$ was not determined by $V_1$ and $V_2$ components alone, but there was an independent schwa component contributing to the predictability of the trajectory. This target was the mean tongue position for all the full vowels.[3]

Figure 3.3, on the other hand, illustrates the spectral patterns taken at the midpoint of the first /ɪ/, /ə/ and the second /ɪ/ in the sequence /ɪbəbɪ/ in the sentence 'The campaign for Women's **Lib aby**smally failed' spoken by an English

[3]In Browman & Goldstein (1992c), the mean targets for contextual vowels were determined by the extremum in the movement of the two pellets, M, placed at the middle of the tongue dorsum, and R, placed at the rear of the tongue dorsum. Both horizontal (X) and vertical (Y) movements were measured. Thus, each vowel was characterized by four positions, MX, MY, RX, and RY.

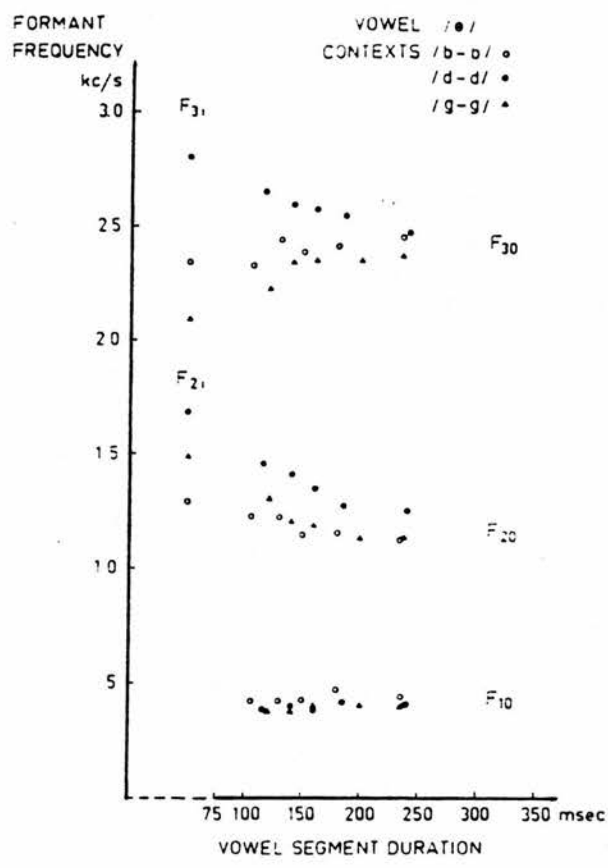



FIG. 2. The first three formant frequencies of the vowel /ə/ are shown as a function of vowel-segment duration and consonantal context. For $F_{10}$, $F_{20}$, and $F_{30}$, each datum point represents the mean of 4 individual measurements adjacent along the duration axis. $F_{2i}$ and $F_{3i}$ data refer to the mean of approximately 20 measurements.

**Figure 3.1.** Figure taken from Lindblom (1963:p1775).

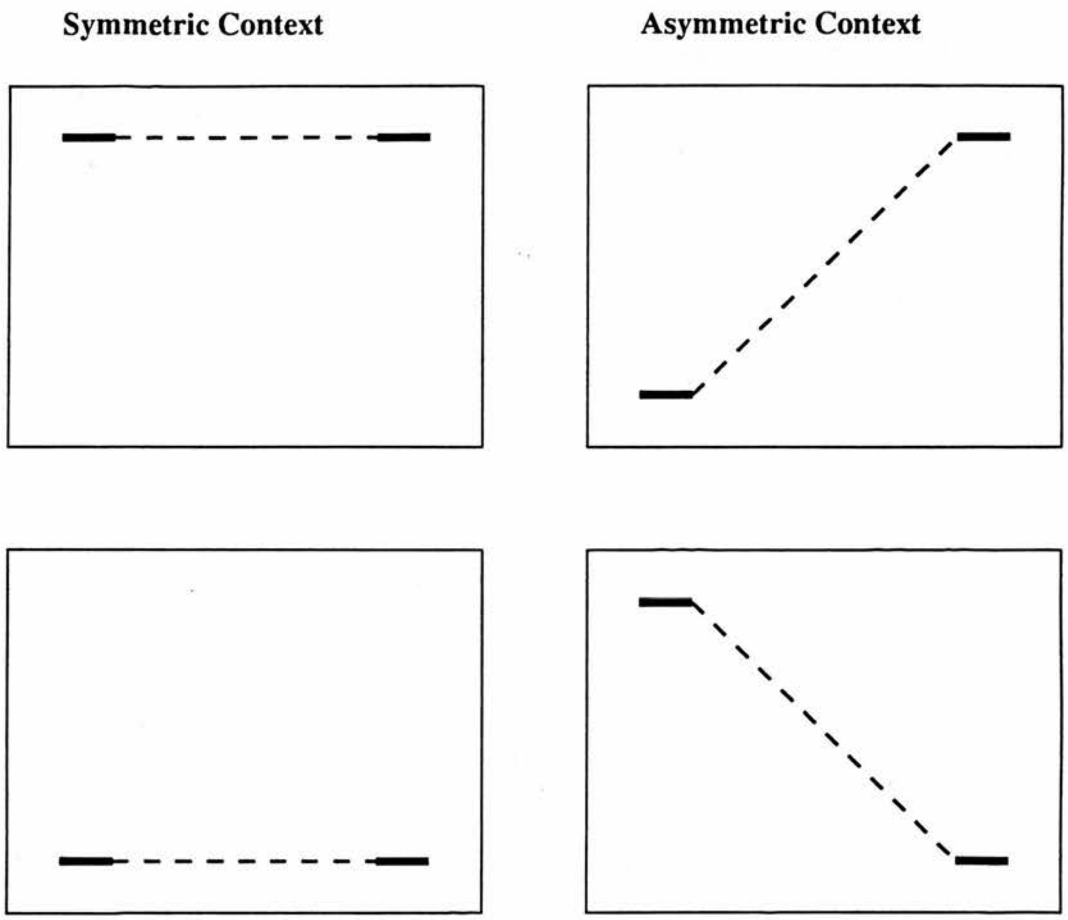



**Figure 3.2.** Schematic illustrations of trajectories across an unspecified segment.

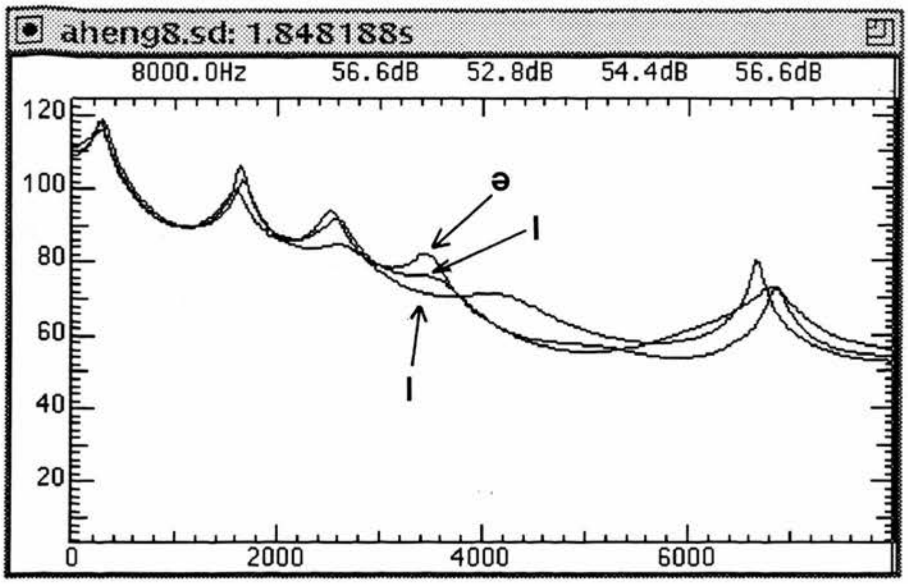



**Figure 3.3.** The spectral patterns taken at the midpoint of the first /ɪ/, /ə/ and the second /ɪ/ in the sequence *Lib abysmal* produced by Subject AH. The spectral pattern for the first /ɪ/ has higher $F_4$ than the other two.

speaker (AH). Though schwa is between two like vowels, the spectral peaks for the three vowels more or less overlap with one another. Figure 3.4 shows his $F_1$ and $F_2$ trajectories for the sequences /ɪbəbɪ/, /ɪbəbæ/, /æbəbɪ/, and /æbəbæ/. His schwas are more or less interpolated through by the $F_2$ trajectories as predicted by Figure 3.2. On the other hand, the $F_2$ trajectory of the sequence /ɪbəbɪ/ produced by another English speaker (DG) (Figure 3.5) shows a dip during the middle schwa.

Here are two pieces of evidence, one of which suggests that schwa is targetless while the other suggests that schwa is targeted. Is schwa targeted or targetless? Or are coarticulatory strategies idiosyncratic, as suggested by Nolan (1985)?

Another important question is how schwa is affected by its immediately adjacent consonants. In the articulatory study of Browman & Goldstein (1992c), consonantal effects on schwa production were not studied. However, Recasens (1986) observed great effects of consonantal contexts on schwa in Catalan. Bergem

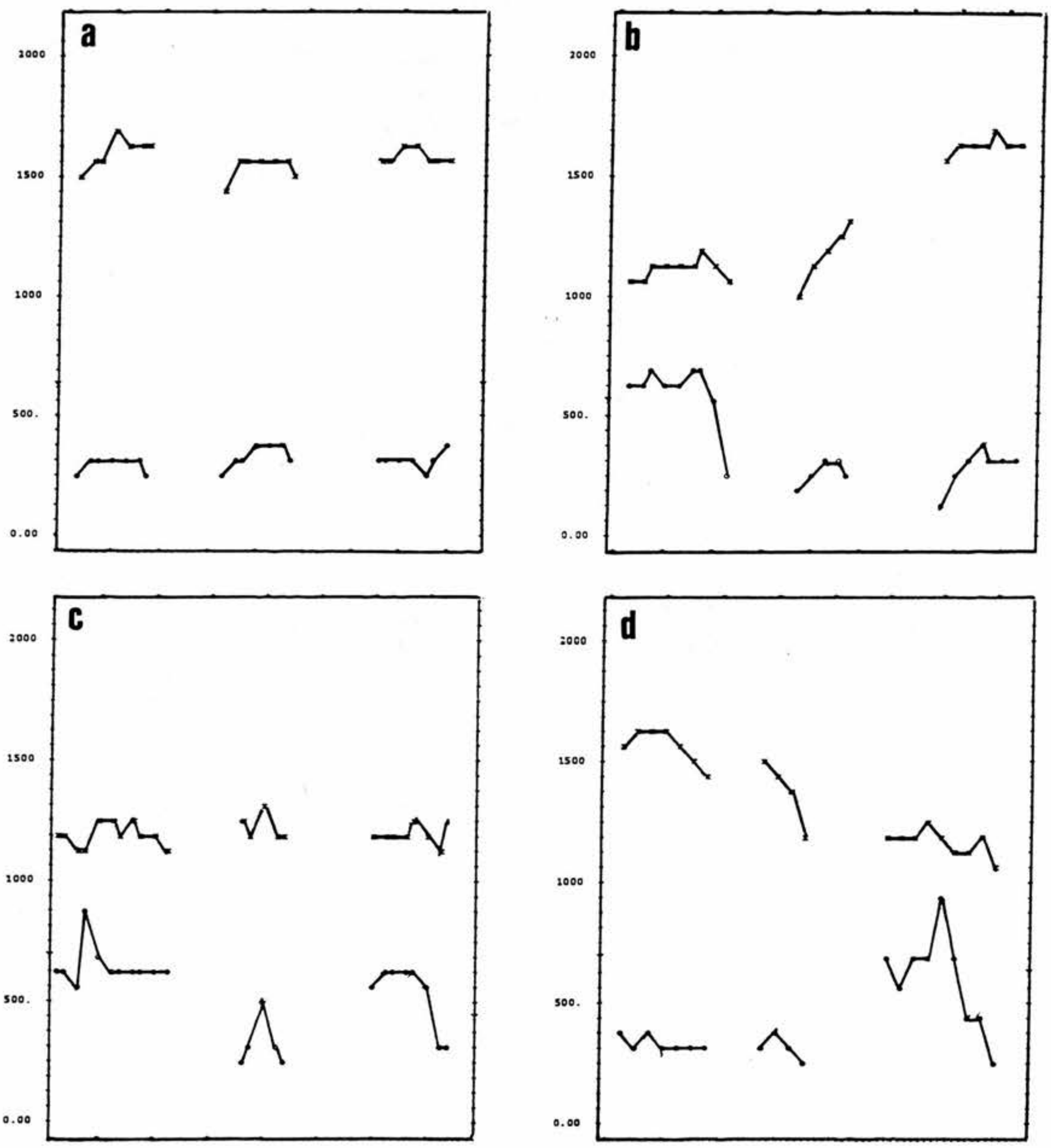



**Figure 3.4.** $F_1$ and $F_2$ trajectories for the sequences (a) /ɪbəbɪ/, (b) /æbəbɪ/, (c) /æbəbæ/, and (d) /ɪbəbæ/ for Subject AH. These are representative samples. The LPC analysis window moved in 10 ms steps.

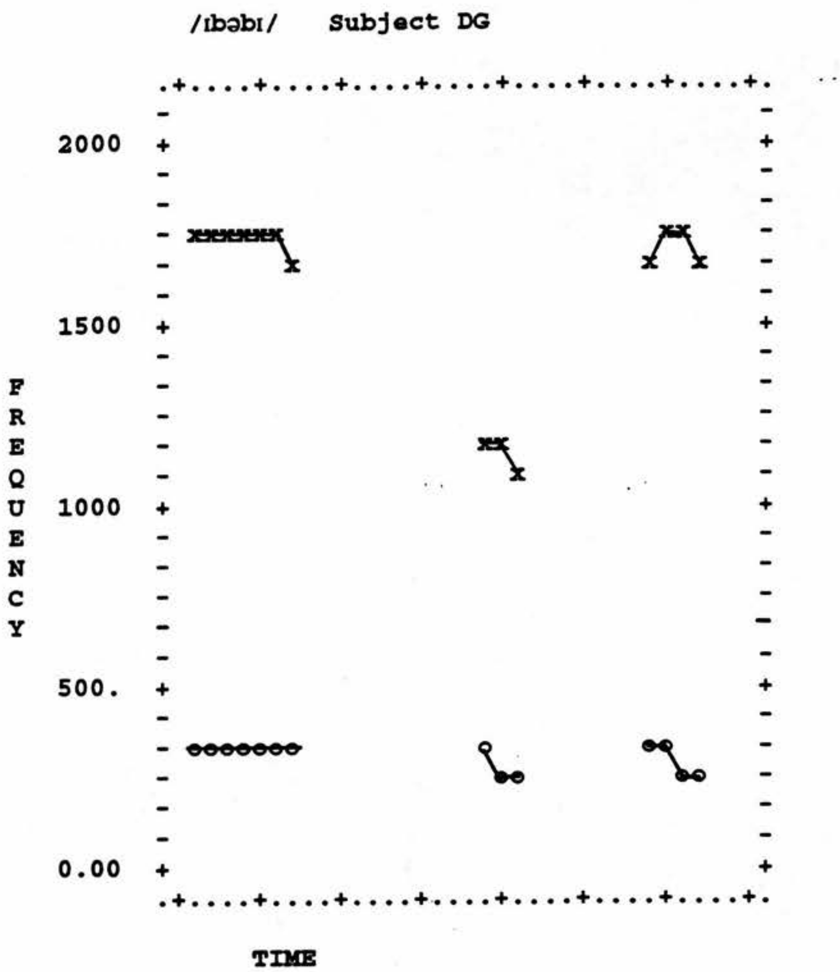



**Figure 3.5.** The $F_1$ and $F_2$ trajectories of /ɪbəbɪ/ of Subject DG. The LPC analysis window moved in 10 ms steps.

(1993) also reports that formant values of a full vowel /ɛ/ and schwa in Dutch shifted to an /ɔ/-like position in the vowel plane in the consonantal context of *w_l*. Schwa resulted in a more extreme position, 346 Hz for $F_1$ and 940 Hz for $F_2$, than the vowel /ɛ/ (393/1146 Hz). Consonants are known to perturb the $F_2$ value of a vowel towards a more central value (Stevens & House 1963). Consonants have also been observed to exert different degrees of constraint on the adjacent vowels and, at the same time, to reduce or block coarticulatory effects from the vowels (Bladon & Al-Bamerni 1976; Farnetani 1990; Recasens 1991; Farnetani & Recasens 1993). How then is schwa affected by the neighbouring consonants? Is the V-ə-V trajectory perturbed by the intervening consonants? Does a consonant reduce or block vowel-to-schwa coarticulation? Could the dip observed during schwa for the speaker DG (Figure 3.5) be attributed to consonantal effect while AH seems to show less C-to-V constraint (or more V-to-C effects)?

## 3.2 Experiment 1

The present study sets out to test the hypothesis above. That is, schwa is a product of contextual assimilation and it has no target value of its own. An experiment was designed to study the acoustic variability of schwa as a function of context using sentences as natural as possible. Nonsense utterances were avoided by placing VCəCV sequences within natural sentences.[4] The effects of C-to-V as well as V-to-V coarticulation on schwa were studied using the symmetric VCəCV sequences with the consonants /p, t, k/ and the vowels /ɪ, æ, u/.

[4]There are drawbacks in placing VCəCV sequences within natural sentences. For example, in the present study, it was impossible to control prosodic factors. Thus, different types of boundaries (both prosodic and syntactic) intervene within the VCəCV sequence, and sometimes a long pause is observed between the preceding context and the schwa. The place of the nuclear accent also moves in relation to the VCəCV sequence from one sentence to another. Such uncontrolled variables may have affected the results of the present experiment. However, it was decided that nonsense words should be avoided. This is because the present material will also be used to test the performance of non-native speakers in the later experiment. It was considered that the production of nonsense words in L2 may introduce unknown factors that might affect the results of the experiment. The material used in the present experiment is a compromise between a good control of segmental contexts and more natural production in an experimental setup.

The hypothesis to be tested is: schwa is phonetically unspecified with no target value of its own; its acoustic values and trajectories are determined solely by its contexts. The hypothesis will be tested as follows. If the acoustic value of schwa is determined by context alone, the schwa trajectory from the preceding segment to the following segment will be level in the symmetric context without passing through any inherent target of its own. Then there will be a perfect correlation between the schwa onset ($x$) and the schwa midpoint ($y$) value as well as between the schwa offset ($x$) and the midpoint ($y$) value with no $y$-intercept, i.e., $y = x$. That is, though some variability in production is expected, if the prediction of the hypothesis is to bear out, the $y$-intercept $\beta_0$ must be close to zero, while the regression coefficient $\beta_1$ must be close to 1.0 in the linear regression equation $y = \beta_0+\beta_1 x$. Regression analyses were performed to test this hypothesis.

The labial, alveolar and velar consonants are expected to affect the $F_2$ values of schwa at the transitions. Thus, a wide range of $F_2$ values are expected at the schwa onset and offset as a function of consonantal contexts. If $F_2$ trajectories across schwa are level despite this wide range, it will strongly support the targetlessness of schwa.

Multiple regression analyses were also performed to see how much of the contextual effect may be proportioned to the consonant and vowel context. The effects of the transconsonantal vowels on the consonantal loci were also observed.

The directionality of coarticulatory influence were also studied. A number of conflicting results have been reported on the relative strength of carryover (L-to-R) and anticipatory (R-to-L) effects in English (Sharf & Ohde 1981). The results of C-to-V and V-to-C perceptual studies suggest that anticipatory effects are stronger than carryover effects in English (Ohde & Sharf 1977; Ostreicher & Sharf 1976; Sharf & Beiter 1974; Sharf & Hemeyer 1972; Winitz *et al.* 1972). On the other hand, the results of acoustic and articulatory studies ($F_2$) seem to suggest that carryover effects are greater (Gay 1974; Bell-Berti & Harris 1976; Ohde & Sharf 1975; Öhman 1966). The results of the present experiment will be later compared with the coarticulatory pattern observed for the Japanese vowels.

The $F_1$ and $F_2$ values of schwa were measured and their trajectories across time (the onset, midpoint, and offset of a schwa) were observed in different vocalic and consonantal contexts. The first two formants were studied, as they are known to correlate with tongue height and backness movement in articulation. Schwa may be targetless in both formants, that is, both in [Height] and [Backness] features. It is also possible that schwa may be targetless in one, but targeted in the other feature.

### 3.2.1 Methods

#### Materials

VCəCV sequences with possible combinations of the three vowels /ɪ, æ, u/ and the three consonants /p, t, k/ were studied. All the contexts were symmetrical, i.e., $V_1$ and $V_2$ were identical, as well as $C_1$ and $C_2$. This resulted in 9 possible sequences as follows.

1. ɪpəpɪ: Please dip **a** pin in the solution.
2. æpəpæ: Please wrap **a** package with this paper.
3. upəpu: Through the hoop **a** poodle jumped
4. ɪtətɪ: They are going to fit **a** timber roof on our house.
5. ætətæ: We gave the cat **a** tag to wear round its neck.
6. utətu: The robbers went to loot **a** tomb.
7. ɪkəkɪ: You may pick **a** kitten from the basket.
8. ækəkæ: I forgot to pack **a** can of coffee for him.
9. ukəku: I gave Luke **a** cool drink.

Each sentence was repeated 10 times by each speaker in a randomized order. The schwa in the function word *a* is studied in the present experiment.

### Speakers

Three male speakers, AH, MB, and DG participated in the experiment. They were MSc. students in the Department of Linguistics, University of Edinburgh, at the time of the recording. AH, originally from Manchester, speaks standard North Western English. MB, from Lincolnshire, and DG, from the south of England, speak standard Southern British. All three speak non-rhotic English.

### Recordings

The recording was done in a sound-treated recording studio in the Department of Linguistics, Edinburgh University. The speakers read a list of randomized sentences with various VCəCV sequences described above.

### Analyses

The sentences for analysis were sampled at 16 kHz into a UNIX SUN workstation with WAVES speech analysis facilities. The VCəCV sequences to be studied were hand-labelled by using the XLABEL system of WAVES, displaying the waveform and a wide band spectrogram. Formant values were obtained by running the FORMANT program for LPC analysis with a 25 ms cos**4 window moving in 5 ms steps. FORMANT creates a file which gives the first four formant values and their bandwidths. This was aligned with the label file. When the middle of the analysis window fell within the boundaries of a segment, the record of the formant values obtained by the window was assigned to the segment. When the middle of the window fell at the initial boundary of a segment, the record was assigned to this segment. On the other hand, when the middle of the window fell at the final boundary, this record was assigned to the next segment. In this way files for all the $V_1$, ə, and $V_2$ segments with the first four formants and their bandwidths at 5 ms intervals were obtained. In order to observe formant trajectories, from the onset transition to the offset transition of a schwa, the first, middle and the last formant frequency values of a schwa segment were recorded. The LPC analysis was run by moving the window in 5 ms rather than 10 ms steps to obtain enough record numbers as schwa segments are short and sometimes have only a couple of glottal pulses. Where there were even numbers of records for a segment, the mean of the middle two values was chosen as the midpoint value.

Formant tracking errors were corrected where possible. Otherwise, values judged spurious were not used and were treated as missing values. Formant tracks were overlaid on top of spectrograms to check errors. Many of the errors resulted from the confusion between different formants and correction was possible. Most of the $F_2$ values were recovered this way. Hand correction was possible, but this was avoided as systematic error could be introduced by the handling of the cursor in deciding the formant value. In segmenting, the first and last weak waves at the edges of a segment were not included, to avoid spurious formant values.

By running the STATS program, the mean, maximum and minimum formant values of $F_1$ and $F_2$ for each segment were also calculated. For this purpose, the LPC analysis was run by moving the window in 10 ms steps.[5]

#### Statistics

The BMDP statistical package was used to run analyses of variance (ANOVAs) and to run regression analysis. Scatterplots used in the figures were also obtained by the BMDP program. Statistical analyses were performed using the vowel midpoint frequency values. ANOVAs were performed for each speaker where there were enough observations per speaker. Otherwise, pooled data across speakers were used. Post hoc scheffe tests were also performed to test which differences between means are responsible for a significant effect in an ANOVA.

### 3.2.2 Results

Two-way ANOVAs were performed with consonantal and vocalic contexts as grouping factors for each speaker. Table 3.1 shows the level of significance for

[5]One-way ANOVAs were performed between the mean values and the midpoint values of schwa $F_1$ and $F_2$ across speakers. There was a significant difference due to different measurements in $F_1$: $F(1,496) = 4.83$, $p = 0.0285$. The $F_1$ values were generally higher at the vowel midpoint (294 Hz) than for the vowel mean (282 Hz). The standard deviation was also higher at the vowel midpoint (73 Hz) than for the vowel mean (47 Hz). However, no significant difference was observed for $F_2$: $F(1,517) = 0.04$, $p = 0.8337$. The mean values biases the results in favour of the 'targetlessness' as the mean value across a segment includes more contextual information. In order to avoid such a bias, the vowel midpoint values were used for analysis.

| | p-value | | | | | |
|---|---|---|---|---|---|---|
| | $F_1$ | | | $F_2$ | | |
| Subj. | $C$ | $V$ | $C \times V$ | $C$ | $V$ | $C \times V$ |
| AH | 0.0226 | 0.2002 | 0.9056 | 0.0000 | 0.0000 | 0.0000 |
| MB | 0.0001 | 0.7812 | 0.3215 | 0.0000 | 0.0000 | 0.0010 |
| DG | 0.0451 | 0.4602 | 0.7869 | 0.0000 | 0.0000 | 0.0000 |

**Table 3.1.** The level of significance of the results of ANOVAs. The table shows the main effects of consonant and vowel contexts and their interaction on the first and second formants of schwa for the three speakers.

the two main effects and the group interaction on the first and second formants of schwa for each speaker.

The main effects of both consonantal and vocalic contexts were significant for $F_2$ ($p < 0.001$) for all of the speakers. The main effects of consonantal context were significant for $F_1$ ($p < 0.05$). No significant effects of vowel-to-schwa coarticulation was observed in $F_1$ for any of the speakers. There was no significant interaction of consonantal and vocalic contexts in $F_1$. For $F_2$, the interaction was significant. The effects of the vocalic context seem to vary as a function of consonantal context and vice versa. In particular, the assimilatory effect of a velar consonant /k/ with the neighbouring vowels affects the coarticulatory pattern of schwa in the VCəCV sequences as will be described below.

Table 3.2 shows the matrices for each speaker of the mean $F_1$ and $F_2$ values of schwa in the 3 consonant $\times$ 3 vowel contexts. The rightmost value on the bottom row is the grand mean for each matrix.

### $F_1$

There were no significant effects of vocalic contexts on the $F_1$ values of schwa. The $F_1$ values of schwa are higher in the context of the labial consonant /p/ than in those of alveolar or velar consonants. A careful observation of the vertical movement of a pellet placed at the rear of the tongue dorsum in Browman & Goldstein (1992c) shows that there is a dip in the trajectory related to the bilabial consonant /p/ at the rear of the tongue. Such slight lowering of the rear of the

| $F_1$ | | | | | $F_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| AH | p | t | k | | AH | p | t | k | |
| ɪ | 268 | 233 | 251 | 252 | ɪ | 1504 | 1592 | 1964 | 1697 |
| æ | 295 | 259 | 262 | 270 | æ | 1308 | 1517 | 1698 | 1508 |
| u | 272 | 249 | 264 | 262 | u | 1372 | 1644 | 1494 | 1503 |
| | 277 | 248 | 259 | 261 | | 1394 | 1583 | 1718 | 1565 |
| MB | p | t | k | | MB | p | t | k | |
| ɪ | 344 | 321 | 307 | 324 | ɪ | 1494 | 1787 | 2044 | 1775 |
| æ | 350 | 305 | 322 | 323 | æ | 1370 | 1687 | 1988 | 1681 |
| u | 362 | 303 | 291 | 317 | u | 1275 | 1732 | 1737 | 1581 |
| | 352 | 310 | 307 | 321 | | 1380 | 1735 | 1923 | 1679 |
| DG | p | t | k | | DG | p | t | k | |
| ɪ | 286 | 265 | 295 | 282 | ɪ | 1123 | 1569 | 1908 | 1562 |
| æ | 315 | 286 | 294 | 293 | æ | 1121 | 1498 | 1863 | 1494 |
| u | 306 | 273 | 303 | 293 | u | 1129 | 1542 | 1456 | 1384 |
| | 300 | 274 | 297 | 289 | | 1124 | 1535 | 1742 | 1478 |

**Table 3.2.** The $F_1$ and $F_2$ values at the midpoint of schwa in the 3 consonant /p, t, k/ × 3 vowel /ɪ, æ, u/ contexts for each speaker. The rightmost column of each matrix shows the mean formant frequencies for the vocalic contexts. The bottom row shows the mean formant frequencies for the consonantal contexts. The rightmost value on the bottom row is the grand mean for each matrix.

tongue may be a possible source of schwa's higher $F_1$ value in the context of /p/. Engstrand (1983) also reports that the tongue interacts with a V-to-V gesture during the labial consonant /p/ in Swedish. The tongue is lowered during the production of /p/ in VCV articulation. There may be two reasons for this. First, the CV transition of a voiceless plosive is typically characterized by the following five phases: (1) occlusion (2) transient (3) frication (4) aspiration and (5) the onset of the periodic waves (Fant 1973). To maximize the acoustic effect of the CV transition, the speaker controls the timing of the vowel-related tongue gesture in such a way that supraglottal turbulence noise is avoided during the interval. This requires a relatively lowered tongue position. Secondly, a relatively wide supralaryngeal cavity would be favourable to the labial /p/ burst since the pressure-drop caused by a narrow constriction behind the lips would reduce the noise generating efficiency at the primary constriction. Differences in the mean $F_1$ values as a function of consonantal contexts are however small. They are 18 Hz, 45 Hz and 3 Hz for AH, MB and DG.[6]

Though some systematic variability in $F_1$ as a function of context was observed across the three speakers, differences in $F_1$ values as a function of context are very small. The $r^2$ values for the onset/midpoint and offset/midpoint correlation for $F_1$ are also small (Table 3.3). Context dependent systematic variability is generally small for $F_1$. On the other hand, the means of differences between the maximum and minimum values of $F_1$ within a segment are relatively large; 82 Hz, 93 Hz and 60 Hz for AH, MB and DG, suggesting that $F_1$ trajectories during schwa do not stay level between the surrounding VC and CV which are identical in segmental makeup. Figure 3.6 shows the mean $F_1$ trajectories through the onset to the offset of schwa as a function of contexts across three speakers.[7] These trajectories in general show a rise at the midpoint. However, although they don't stay level, they don't seem to deviate towards a common target, either. These results seem to suggest that schwa may not be unspecified in $F_1$, but from the present data

[6]The $F_1$ values of the vowel /a/ in Japanese were also observed to be higher in the labial /p/ context than in the alveolar or velar contexts (see page 193).

[7]Each point in the graph is the mean of 10 repetitions × 3 speakers. In some cases there are less than 30 observations per mean due to formant tracking errors. Also, some schwas were devoiced, extremely shortened or deleted and formant values were not obtained. The smallest observations per mean is 15 in the ap_pa context for $F_1$ and 26 for $F_2$.

we may not conclude that schwa is targeted, either.

### $F_2$

The extent of variation observed in the second formant of schwa due to consonantal contexts was large and systematic across speakers. The mean $F_2$ values increased according to context, p < t < k. Recasens (1986) also observed that $F_2$ values of Catalan schwa increased according to context, p < t < k. Vocalic effects on $F_2$ were lesser in degree compared to the effects of consonants, but they were systematic. The mean $F_2$ values of schwa increased according to context, u < a < i.

The interaction of C-to-V and V-to-V effects on schwa was significant for $F_2$. In the contexts of the front vowels /ɪ/ and /æ/, the mean $F_2$ values increased according to context, p < t < k while in the context of the back vowel /u/ the $F_2$ values increased according to context, p < k < t for AH and DG. For MB, when the transconsonantal vowel was /u/, the difference in $F_2$ value between schwas in the context of /t/ and those in the context of /k/ became negligible. The $F_2$ values of schwa also showed greater spread as a function of transconsonantal vowels in the context of /k/.

It is generally observed that velar consonants assimilate with the neighbouring vowels, especially with the following vowel (Gimson 1980; Öhman 1966), in the place of articulation. Thus, in the context of a front vowel, the place of articulation for /k/ is fronted (*kitten*), while in the context of a back vowel, its place of articulation is retracted (*cool*). As a result, there are different places of articulation for /k/, depending on its vocalic contexts. This, in turn, seems to affect schwa. That is, schwa in the context of a fronted /k/ has a higher $F_2$ value, whereas schwa in the context of a back /k/ has a lower $F_2$ value. In this experiment, as the $F_2$ value of the back vowel /u/ was lowered before a velarized /l/ in the word *cool*, the contrast between a fronted and a retracted /k/ was further enhanced. (There will be more discussion on the interaction of V-to-V and C-to-V effects below (page 59).)

Table 3.3 shows the correlation between the onset/offset formant values and the formant values at the midpoint of schwa for each speaker. The schwa onset

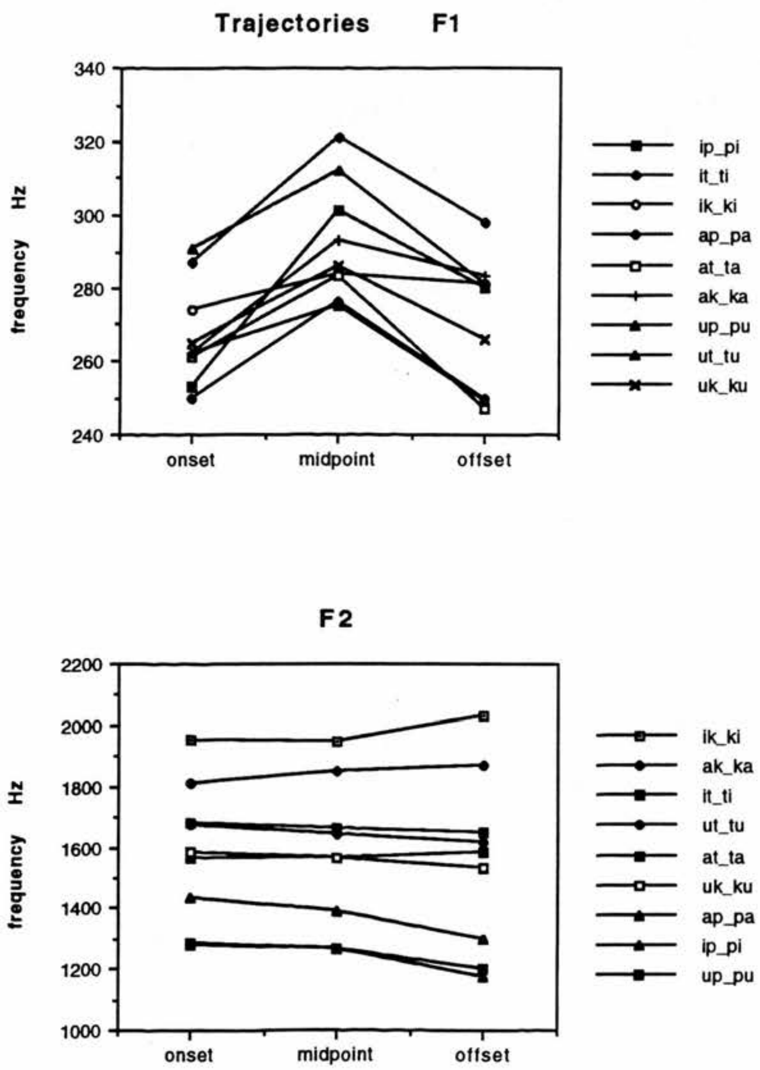



**Figure 3.6.** The mean $F_1$ and $F_2$ values of schwa at the initial transition, midpoint and the final transition across speakers in the contexts of /p, t, k/ and /ɪ, æ, u/. Each point is the mean of 10 repetitions × 3 speakers. In some cases there are less than 30 observations for each point due to schwa devoicing, deletion and formant tracking errors.

| | | $F_1$ | | $F_2$ | |
|---|---|---|---|---|---|
| | Subj. | onset | offset | onset | offset |
| $R^2$ | AH | 0.2671 | 0.1126 | 0.8871 | 0.7745 |
| | MB | 0.0027 | 0.0645 | 0.9537 | 0.9169 |
| | DG | 0.1721 | 0.1293 | 0.9272 | 0.9215 |

**Table 3.3.** The results of the regression analyses for the variability in the formant values of schwa. The independent variables are the onset and offset formant values of schwa. The dependent variable is the formant value at the midpoint of schwa.

(the CV transition) and offset (the VC transition) values represent the preceding and the following consonantal contexts respectively. Though significantly correlated, the r square values are small for $F_1$, while they are mostly over 0.9 for $F_2$. The linear regression equation for the schwa onset ($x$) and the schwa midpoint ($y$) across speakers is $y = 1.0011x - 13.05$. The regression coefficient is 1.0011, which is very close to 1, and the $y$-intercept is $-13.05$, which is very small. This equation suggests that schwa may be targetless in $F_2$. On the other hand, the offset/midpoint correlation is not as good. The onset $F_2$ value is more strongly correlated with the midpoint value of schwa than the offset $F_2$ value for all the speakers. The linear regression equation for the schwa offset ($x$) and the midpoint ($y$) is $y = 0.7800x + 367.98$. The divergence of the $F_2$ trajectories towards the offset due to asymmetry in the strength of contextual effects seems to explain this equation. Figures 3.7 and 3.8 show the correlation between the onset/midpoint and offset/midpoint $F_2$ values respectively.

Figure 3.6 shows the mean $F_2$ trajectories of schwa in the 9 VCəCV sequences across three speakers. The extent of $F_2$ variability as a function of contexts is observed not only at the transitions but also in the middle of schwa. The range observed in Figure 3.6 is about 700 Hz at the vowel midpoint. The mean schwa duration for these trajectories is 34 ms. $F_2$ trajectories across schwa between identical segments seem to be more or less level, especially from the onset to the midpoint. Trajectories seem to diverge towards different points at the offset though the context is symmetrical. A velar /k/ next to a front vowel /ɪ/ is fronted and has higher $F_2$, while a velar next to a back vowel /u/ is retracted and

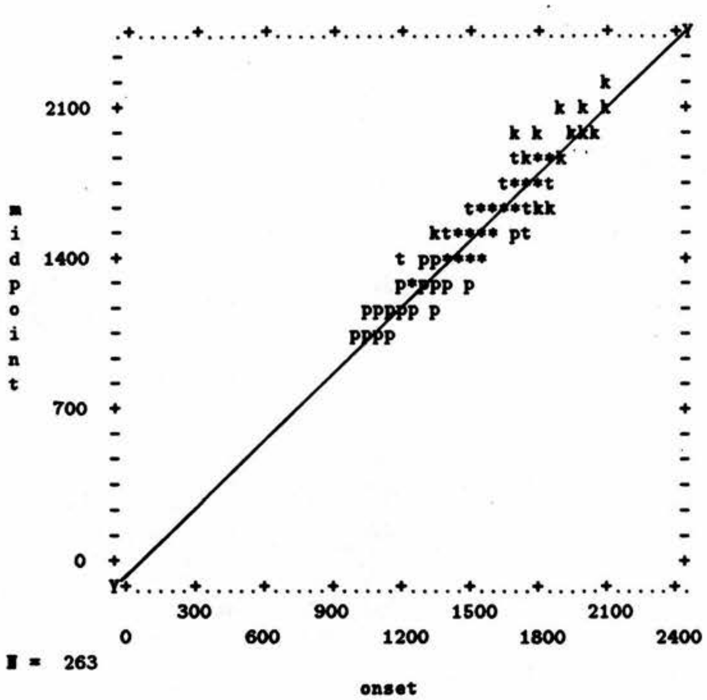



**Figure 3.7.** The correlation between the onset and midpoint $F_2$ values.

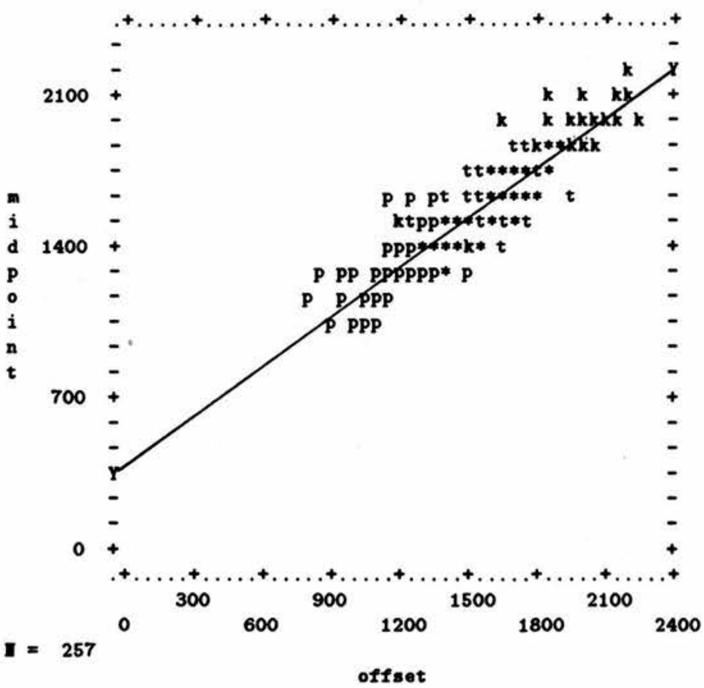



**Figure 3.8.** The correlation between the offset and midpoint $F_2$ values.

has lower $F_2$. This effect was stronger at the schwa offset as the following vowel exerts stronger effect on /k/ than the preceding vowel. Thus, vowel-to-vowel effects were stronger at the offset in the context of the velar /k/. Consonant-to-vowel effects also seem to be stronger at the offset for all the speakers. The mean differences as a function of consonantal contexts (/p/ vs. /k/ for the front vowel context and /p/ vs. /t/ for the back vowel context) were generally greater at the vowel offset than at the onset (Table 3.7). There seems to be asymmetry in C-to-V effects between the onset and offset of schwa. This asymmetry in the C-to-V effects resulted in a slightly less steep slope with larger $y$-intercept for the offset/midpoint regression compared to the onset/midpoint regression equation which shows nearly perfect correlation between $x$ and $y$ with the slope of about 45°.

The above seems to suggest that consonantal loci may be different at the initial and the final transition of schwa. Variables such as the place of the nuclear stress and segments beyond the VC_CV sequence, that are not controlled in this experiment, may also be affecting the pattern of the trajectories. Choi (1992) also reports asymmetry in consonantal loci at the onset and offset of the vowel in symmetric CVC syllables in Marshallese. The offset $F_2$ values were lower than the onset $F_2$ values. Choi reports the differences of 179 Hz for the alveolar context and 197 Hz for the labial context. The situation in Marshallese is slightly complicated as both $C_1$ and $C_2$ are characterized by secondary as well as primary articulation. For the primary articulation, Kent & Moll (1972) and Munhall *et al.* (1991) report similar asymmetries in articulatory studies. They have shown that the tongue body moves in an elliptical path out of and into the consonants in CVC syllables. Kent & Moll (1972) suggest that the asymmetries are due to aerodynamic factors. There is a greater build-up in air pressure in the production of the syllable-initial consonant, implying a forward movement at release. In the production of the syllable-final consonant, the pressure in the oral cavity has equalized to atmosphere during vowel production. Thus, the constriction during the final consonant would not be subject to the same aerodynamic pressures as the initial consonant. In the VCəCV sequence in the present study, $C_1$ is ambisyllabic in nature and thus seems to be syllable-initial to the schwa. In the labial and alveolar contexts, $F_2$ trajectories are declining. The differences are

| | AH | MB | DG |
|---|---|---|---|
| | $R^2$ | $R^2$ | $R^2$ |
| C | 0.5262 | 0.7415 | 0.6386 |
| $V_1$ | 0.2643 | 0.1487 | 0.2019 |
| $V_2$ | 0.1908 | 0.0415 | 0.0255 |
| $V_1 + V_2$ | 0.3265 | 0.1598 | 0.2073 |
| $C + V_1$ | 0.6565 | 0.7557 | 0.7064 |
| $C + V_2$ | 0.8641 | 0.8928 | 0.8671 |
| $C + V_1 + V_2$ | 0.8742 | 0.8932 | 0.8712 |

**Table 3.4.** The results of the multiple regression analyses for the $F_2$ variability of schwa. The independent variables are consonantal context (C), $F_2$ value of the preceding vowel ($V_1$) and the following vowel ($V_2$). The $F_2$ measurement for schwa is taken at the vowel midpoint. The $F_2$ value for the contextual vowel was also measured at the vowel midpoint. For consonantal contexts, arbitrary scores of 1 to 3 were given for /p, t, k/.

110 Hz in the labial context and 22 Hz in the alveolar context. The negative $y$-intercept for the onset/midpoint correlation also matches the account.

Multiple regression analyses were performed to see the predictability of schwa's midpoint $F_2$ values as a function of consonantal and vocalic contexts. In the multiple linear regression model, the value of the dependent variable $y$ is predicted by the equation, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$, where $x_i$ is the $i$ th independent variable, $\beta_i$ is the $_i$ th regression coefficient, $\beta_0$ is the intercept, p is the number of independent variables. The analyses were performed for each speaker. For the consonantal contexts, arbitrary scores of 1 to 3 were given for /p, t, k/. For the vocalic contexts, the $F_2$ values at the midpoint of the preceding and the following vowel were entered separately. It must be noted that in this experiment, $V_1$ and $V_2$ were always identical. Thus, a strong correlation was observed between the preceding and the following vowel. Table 3.4 summarizes the results. The table shows that for example, for AH, the consonant context accounts for 52% of the total $F_2$ variability observed at the midpoint of schwa, while $V_1$ and $V_2$ independently account for 26% and 19% of the total variance of schwa in $F_2$ respectively. The combination $C + V_1 + V_2$ accounts for 87% of the total variance of schwa.

For all of the subjects, the consonantal context is the most powerful predictor of the schwa $F_2$ value. For MB and DG, consonants alone account for 74% and 63% of the total variability of schwa $F_2$ values. For AH, though consonants still play a major role in determining the $F_2$ value of schwa, they do not place such strong constraints as observed for MB or DG. Among the three speakers, AH showed the least effect of consonantal contexts. Differences in the mean $F_2$ values of schwa in the contexts of /p/ and /k/ are 543 Hz and 618 Hz for MB and DG while it is only 324 Hz for AH. On the other hand, AH showed the strongest vocalic effects. The combination of $V_1$ and $V_2$ accounted for 32.65% of the total variance of schwa $F_2$ values for AH. The preceding vowel is more strongly correlated with the $F_2$ value of schwa than the following vowel for all of the speakers.

All the tokens of $V_1$, schwa and $V_2$ in 90 VCəCV utterances are plotted in Figure 3.9 for each speaker. $V_1$ and $V_2$ are /ɪ, æ, u/. The range of $F_2$ values of schwa seems to cover nearly the whole $F_2$ range of the three contextual vowels. Some tokens of /u/ with low $F_2$ values outside the range of schwas are those whose $F_2$ was pulled down by the following dark /l/ in the word *cool.*[8]

The extent of variation of $F_2$ observed in schwa with 3 consonant $\times$ 3 vowel contexts is large. The ranges from the smallest to the largest value observed for each speaker were 1249-2066 (817) Hz, 1133-2171 (1038) Hz and 1002-2101 (1099) Hz for AH, MB, and DG respectively. These ranges are about 1000 Hz.

Figure 3.10 shows all the tokens of schwa observed for the subject DG. The letters p, t, and k represent the consonantal contexts. Within each consonantal context, transconsonantal vocalic contexts were either /ɪ/, /æ/, or /u/. The extent to which labial consonants pull down the $F_2$ values of schwa is clearly observed in this figure. suggesting that the warp observed in $F_2$ trajectories for DG (Figure 3.5) may be due to the consonantal effect of the labial /b/ rather than to a possibly inherent target of schwa.

Figure 3.11 shows the spectrograms of VCəCV sequences produced by MB. In the sequence /ɪpəpɪ/, the $F_2$ specification for the vowel /ɪ/ is high, but the

[8]This seems to be a case of feature spreading, parallel to the case of a fronted Russian /x/ discussed in Keating (1988) (See page 16 and Figure 2.1). The whole segment of the vowel /u/ seems to have taken on the [Backness] property of the following velarized /l/.

```
       AH                        MB                       DG
       .....+......+.....      ..+....+....+.....      ...+.....+.....+..
       -                -      -                -      -  i             -
  2100 +                +      -                - 2400 +                +
       - x              -      -i               -      -                -
       -  xx            - 2400 +i               +      -                -
       -  x*xi          -      -                -      -                -
       -  x*x  i        -      -                -      -                -
       - iiii  i        -      -                - 2100 + x              +
  1800 +  ii*  iii      +      - x              -      -  *i            -
       -  x*iii    i    -      -ix              -      -  *i            -
       -  ***i          -      -i*i             -      - **ii           -
       -  ***           - 2000 + **i            +      - i**i           -
f      - xx**i          -      -***i            - 1800 + **ii           +
2      -  x**i          -      -i**        a    -      - ***         a  -
  1500 +  ***       a   +      -u**i       aa   -      -  **      aaa   -
       -  x**      aa aa-      -**x      aaaaa  -      -x**u     aaaa   -
       -   *       aaaa -      -x**     aaaaa a -      -x**u  a   aaaa  -
       -  xxx x a  aaa  -      -x*        aaa   - 1500 + **u       aa   +
       -    *x     aaaa - 1600 +x*xx       a    -      - x* x    aaa    -
       -x    x    aa a  -      - ux      a a    -      -  x      a      -
  1200 +                +      -u*xx      aa    -      - xx             -
       -                -      - x        a     -      -   a            -
       -                -      - xxx      aa    - 1200 -   x            +
       -                -      - xxxx  aaa      -      - xxxx           -
       -                -      -  xx  aa        -      - xxx        a   -
       -                - 1200 +     a          +      - xx             -
  900  +                +      - x              -      -                -
       -   u            -      -                - 900  +  u             +
       -   u            -      -                -      -                -
       -   u            -      -uuu             -      - uu             -
       -                -      - uu             -      - u              -
       -                -      -                -      -  u             -
       .....+......+.....      ..+....+....+.....      ...+.....+.....+..
           300                  300       900            300   600   900
                  600                600

           f1
```

**Figure 3.9.** The scatterplots of all the tokens of the contextual vowels /ɪ, æ, u/ and schwa in those contexts for each speaker. Symbols x = schwa, i = ɪ, a = æ, u = u (* means an overlap in value between tokens belonging to different grouping factors).

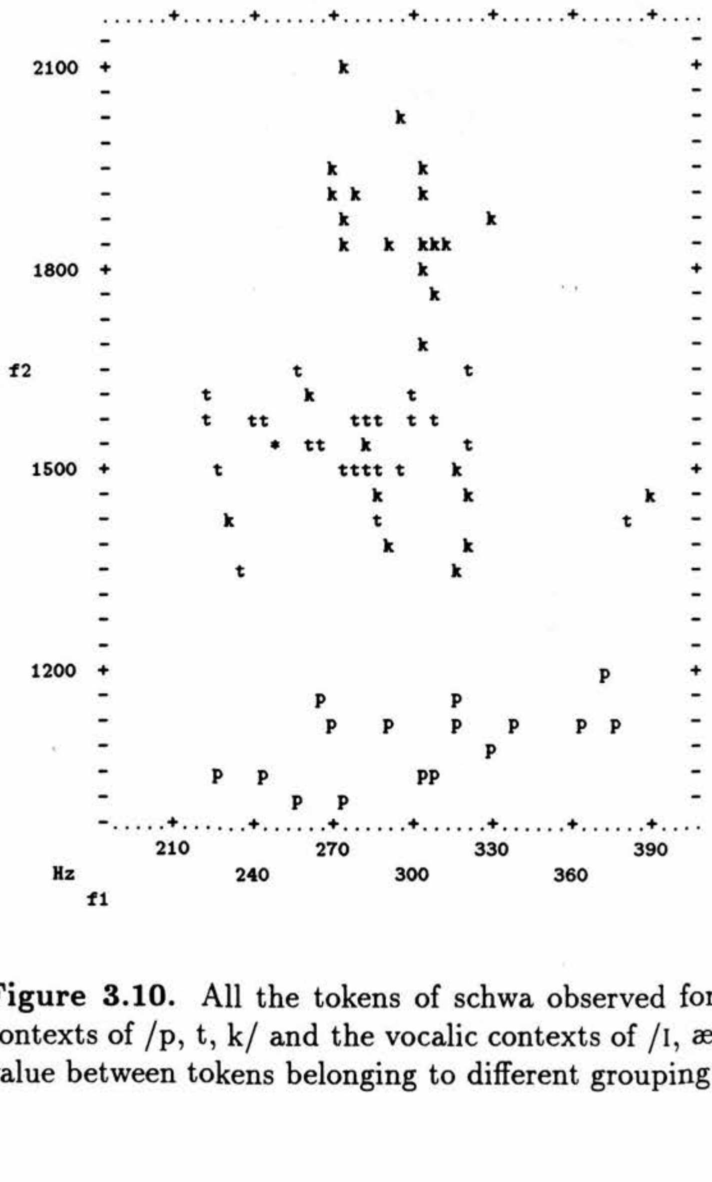



**Figure 3.10.** All the tokens of schwa observed for DG with the consonantal contexts of /p, t, k/ and the vocalic contexts of /ɪ, æ, u/ (* means an overlap in value between tokens belonging to different grouping factors).

specification for the consonant /p/ is low. Thus, the $F_2$ trajectory gradually falls towards /p/ during the first /ɪ/, stays low throughout /ə/, and rises again towards the second /ɪ/ though this rise is masked by the aspiration.

In the case of /ɪkəkɪ/, the $F_2$ specification for the consonant /k/ seems to be even higher than that for the vowel /ɪ/, probably due to a palatalized configuration. Thus, the $F_2$ trajectory rises towards /k/ during the first /ɪ/, stays high during /ə/ and lowers towards the second /ɪ/. Again this fall is masked by the aspiration. In playing the phrase *pick a kitten*, schwa is clearly audible. However, when only the portion of /kək/ is heard, it sounds more like [$c^h$ɪ$c^h$] and schwa on its own sounds more like [ɪ].

For the sequence /ɪtətɪ/, the $F_2$ specification for /t/ seems to be rather high. The $F_2$ trajectory during the two /ɪ/'s stays level. There is a slight dip in the trajectory for schwa. Farnetani (1993) reports that a slight lowering of the tongue body was observed during the sequence /id/, while the tongue tip/blade contact increased from /i/ to /d/ in an EPG study. She suggests that this lowering may have occurred to avoid a strong palatalized configuration behind the front constriction. A slight dip in $F_2$ in the /ɪtətɪ/ sequence may also be attributed to a slight retraction of the tongue body to avoid a palatalized configuration.

These spectrograms illustrate that though the transconsonantal vowel is /ɪ/ in all three cases, the $F_2$ trajectories differ depending on the nature of the intervocalic consonant. Comparing the $F_2$ trajectories, it seems clear that the dip in $F_2$ observed during /ə/ in the sequence /ɪpəpɪ/ is not due to an intrinsic target of schwa, but it is due to the $F_2$ specification for /p/. Otherwise, the $F_2$ trajectory would not have risen towards /k/ during schwa in the /ɪkəkɪ/ sequence.

The range of $F_2$ values of schwa seems to cover the whole range of its contexts, a combination of consonantal and vocalic contexts. The $F_2$ value of schwa is determined mostly by its immediate consonants and, to a lesser degree, by the transconsonantal vowels. In the symmetric VCəCV sequences, $F_2$ trajectories of schwa from the onset to offset are more or less level without going through a target of its own. These results seem to suggest that schwa may be unspecified for $F_2$.

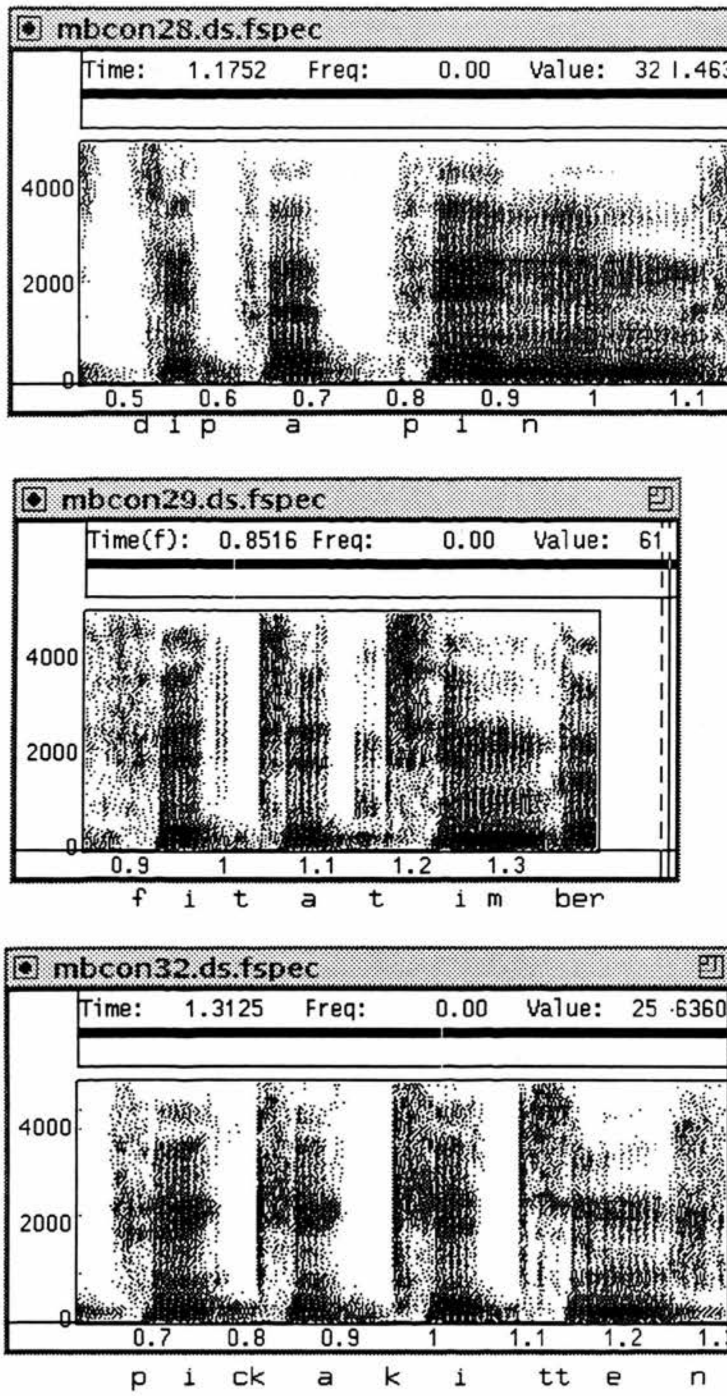



**Figure 3.11.** The spectrograms of VCəCV sequences produced by MB. The utterances are *dip a pin*, *fit a timber* and *pick a kitten*.

| | r | | |
|---|---|---|---|
| | AH | MB | DG |
| $V_1$/onset | 0.6082 | 0.3589 | 0.4748 |
| $V_2$/offset | 0.3916 | 0.1748 | 0.1804 |

**Table 3.5.** The correlation between $V_1$ and the schwa onset $F_2$ values and the correlation between $V_2$ and the schwa offset $F_2$ values.

### V-to-C Effects

At the beginning of this chapter, the question of whether schwa is targeted or targetless was asked, pointing to the fact that the $F_2$ trajectory of schwa in the /ɪbəbɪ/ sequence looked targeted for DG while it looked targetless for AH. It was shown above that schwa may be targetless in $F_2$. That is, its $F_2$ values are determined by context alone. It was also shown that for all the speakers the effect of the consonantal context was greater than the effect of the vocalic context on the $F_2$ value of schwa. However, this does not quite answer the puzzle presented at the beginning of this chapter. When the consonantal context was held constant for /b/, why did AH's trajectories look targetless, while DG's trajectory was seemingly targeted (See Figures 3.4 and 3.5)? This may be due to different degrees of V-to-C effects across speakers. Consonants may themselves be affected by the adjacent vowels in VCəCV sequences. Table 3.5 shows the correlation between $V_1$ and the schwa onset and that between $V_2$ and the schwa offset for each speaker. AH shows the greatest correlation between contextual vowels and the CV/VC transitions of schwa. AH seems to have stronger V-to-C effects than the other speakers. This may suggest that his consonantal loci may vary widely as a function of the transconsonantal vowels, resulting in seemingly targetless trajectories even within a single consonantal context.

### V-to-V versus C-to-V Effects

Another reason why the $F_2$ trajectories of AH and DG looked different may be due to difference in the interaction between the vocalic and consonantal effects across speakers. Table 3.6 shows the mean differences in $F_2$ values as a function of the vocalic contexts /ɪ/ and /u/ in different consonantal contexts for each speaker. In the labial context of /p/, DG shows hardly any difference as a function of vocalic context, while AH and MB show more robust effects.[9] It seems that the labial context blocks the acoustic V-to-V effects in the case of DG. For AH, the labial context does not block V-to-V effects. This difference in the behaviour of the labial consonant seems to explain their apparently different $F_2$ trajectories of schwa in the /ɪbəbɪ/ sequence. Table 3.7 shows the mean differences in $F_2$ values as a function of the consonantal contexts in different vocalic contexts. DG shows greater differences as a function of consonantal contexts than AH does. That is, his C-to-V effects are greater than AH's. However, he shows large V-to-V effects only in the context of /k/. On the other hand, AH shows the least effect of the consonantal context among the three speakers, but his V-to-V and V-to-C effects are greater than the other speakers (See Tables 3.4 and 3.5). The reciprocal relationship observed between the C-to-V and V-to-V coarticulation of DG and AH is suggestive. It seems as though they are compensating for the lack of one with the other. The results of the multiple regression analyses showed very similar $r^2$ values for the combination of $C + V_1 + V_2$ for AH (0.8742), MB (0.8972) and DG (0.8712). Though the magnitudes of V-to-V as well as C-to-V effects may vary from speaker to speaker, in combination they seem to amount to a similar degree. The $F_2$ values of schwa seem to be determined by the combination of the contextual effects of both immediately adjacent consonants and the transconsonantal vowels.

Tables 3.6 and 3.7 show the interaction between the consonantal and vocalic context. The V-to-V effects are greatest in the velar context as discussed above. The /t/ context seems to block V-to-V effects for all the speakers. As the alveolar obstruent /t/ interferes more with the tongue body articulation of a vowel, it may be opaque to V-to-V effects. In the labial context, speakers showed idiosyncratic behaviour. The C-to-V effects are similarly affected by the transconsonantal vocalic context. The difference in $F_2$ value as a function of the consonantal

[9] The level of significance in Tables 3.6 and 3.7 was determined by the post hoc scheffe tests for the interaction of the vocalic context, consonantal context and the points of measurement for each speaker. Because of large differences observed in the /k/ context, some quite robust differences in the /p/ context were not considered significant.

| subject | context | onset | midpoint | offset |
|---|---|---|---|---|
| AH | Vp_pV | 156.6 | 131.9 | 39.5 |
| | Vt_tV | -16.0 | -52.4 | 32.0 |
| | Vk_kV | *427.2 | *470.2 | *542.9 |
| MB | Vp_pV | *246.0 | 219.0 | 197.4 |
| | Vt_tV | 7.6 | 54.2 | 68.5 |
| | Vk_kV | *269.8 | 307.1 | *433.5 |
| DG | Vp_pV | 13.9 | -6.3 | 11.8 |
| | Vt_tV | 15.9 | 26.9 | -9.1 |
| | Vk_kV | *412.0 | *452.7 | *528.4 |

**Table 3.6.** The differences in the mean $F_2$ values of schwa as a function of the contextual vowels /ɪ/ and /u/ (/ɪ/ - /u/) at the different points in the trajectory for different consonantal contexts. The symbol * means that the difference was statistically significant ($p < 0.05$).

context increased according to the vocalic contexts, u < a < i, for AH and DG. This may be explained by the assimilatory effect of the velar /k/ with the surrounding vowels. As /k/ gets fronted in a more fronted context, the acoustic distance between the schwa in the context of /k/ and that in the context of /p/ grows apart. For MB, the acoustic distance is greatest in the /æ/ context. This may be because he has a more palatalized configuration for the /kæ/ than /kɪ/ sequence. In other words, his /k/ may be more fronted in the /æ/ context than in the /ɪ/ context. This seems to suggest that coarticulatory patterns may be highly idiosyncratic and/or dialect specific in fine detail.

### Directionality of Coarticulatory Influence

In Table 3.5 above the correlation between $V_1$ and the schwa onset was greater than the $V_2$/schwa offset correlation for all the speakers. Similarly, the schwa onset/midpoint correlation was stronger than the schwa offset/midpoint correlation (Table 3.3). This seems to suggest that there is asymmetry in the directionality of coarticulatory influence. That is, either carryover or anticipatory is stronger than the other. Table 3.6 shows the differences in the mean $F_2$ values as a function of the vocalic contexts /ɪ/ and /u/ at different points in the schwa trajectory

| subject | context | onset | midpoint | offset |
|---|---|---|---|---|
| AH | iC_Ci | *419.2 | *460.2 | *567.8 |
| | aC_Ca | *345.1 | *390.2 | *456.5 |
| | uC_Cu | *266.7 | *272.4 | *224.4 |
| MB | iC_Ci | *491.0 | *549.2 | *830.6 |
| | aC_Ca | *568.8 | *617.8 | *854.5 |
| | uC_Cu | *492.2 | *457.0 | *560.2 |
| DG | iC_Ci | *695.6 | *785.7 | *851.6 |
| | aC_Ca | *667.7 | *742.4 | *774.5 |
| | uC_Cu | *418.7 | *413.3 | *463.8 |

**Table 3.7.** The differences in the mean $F_2$ values of schwa as a function of the contextual consonants, /p/ and /k/ (/k/ - /p/) for the front vowel contexts and /p/ and /t/ (/t/ - /p/) for the back vowel context, at the different points in the trajectory. All the differences are statistically significant ($p < 0.05$).

for different consonantal contexts. The differences are greatest in the /k/ context for all the speakers, and the differences grow bigger towards the offset. This seems to suggest that anticipatory effects are stronger than carryover effects in V-to-V coarticulation in the /k/ context.[10] In the /p/ context, no significant difference was observed for AH and DG. For MB, significant difference was observed at the onset. Though statistically significant differences were not observed, the differences as a function of the vocalic contexts seem to decrease towards the offset for AH and MB, suggesting that in the /p/ context, carryover V-to-V effects may be stronger for AH and MB. No significant difference as a function of vocalic contexts was observed in the context of /t/ for any of the speakers. Table 3.7 shows the differences in the mean $F_2$ values as a function of the consonantal contexts (/p/ and /k/ for the front vowel context and /p/ and /t/ for the back vowel context) at different points in the trajectory. In general these differences grow bigger towards the schwa offset. It seems that for the C-to-V coarticulation, anticipatory effects are stronger. All the differences were statistically significant ($p < 0.05$).

[10]The reason for this has been described above on page 47.

### V-to-V effects across /b/

The three subjects showed different behaviour for the V-to-V coarticulation across the labial consonant /p/. As discussed above, this seems to explain the seemingly targeted schwa observed for DG in Figure 3.5 and the seemingly targetless schwa observed for AH in Figure 3.4. In this section, V-to-V effects across the labial /b/ are observed more closely with the six vocalic contexts of /ɪ, ɛ, æ, ɑ, ɔ, u/ in order to illustrate different degrees of V-to-V effects across speakers.[11]

The means for $F_1/F_2$ values across all the contexts are 294/1324 Hz, 353/1290 Hz and 319/1177 Hz for AH, MB and DG respectively. The means of the six contextual vowels /ɪ, ɛ, æ, ɑ, ɔ, u/ and the means of schwas in the respective vowel contexts are plotted in the $F_1/F_2$ vowel space for each speaker in Figure 3.12. For AH and MB, the means of schwas in different contexts lie neatly within the six contextual vowels, and systematic relationship between schwa and its contextual vowels may be observed. The effects of transconsonantal vowels are clearly observable for both $F_1$ and $F_2$. For DG, the effects are somewhat less clear.

Figure 3.13 shows the mean $F_2$ trajectories from $V_1$ to $V_2$ across schwa of VbəbV utterances for each speaker. The measurement was taken at the vowel midpoint. Three different degrees of schwa's variability as a function of vocalic contexts may be observed in this figure. This seems to imply that the coarticulatory resistance (see page 23) of the labial /b/ to vocalic effects may vary from

[11]VbəbV sequences with the six contextual vowels /I, ɛ, æ, ɑ, ɔ, u/ were embedded in the following sentences. Only symmetrical cases were studied, i.e., $V_1$ and $V_2$ were always identical. Most of schwas used in this set of data were lexical. Lexical schwa is defined as 'an obligatory phonemic schwa sound' by Koopmans-van Beinum (1994). Each sentence was repeated 10 times in a randomized order. Each speaker read 60 sentences in all.

1. IbəbI: The campaign for Women's Lib abysmally failed.
2. ɛbəbɛ: Yesterday Mr. Webb abetted the stealing of the money.
3. æbəbæ: The crab abandoned its prey as it sensed something approaching.
4. ɑbəbɑ: Next to the patch of rhubarb, a barn stood.
5. ɔbəbɔ: Mission control decided to have the plan to circle the orb aborted.
6. ubəbu: I saw two baboons running away towards the bush.

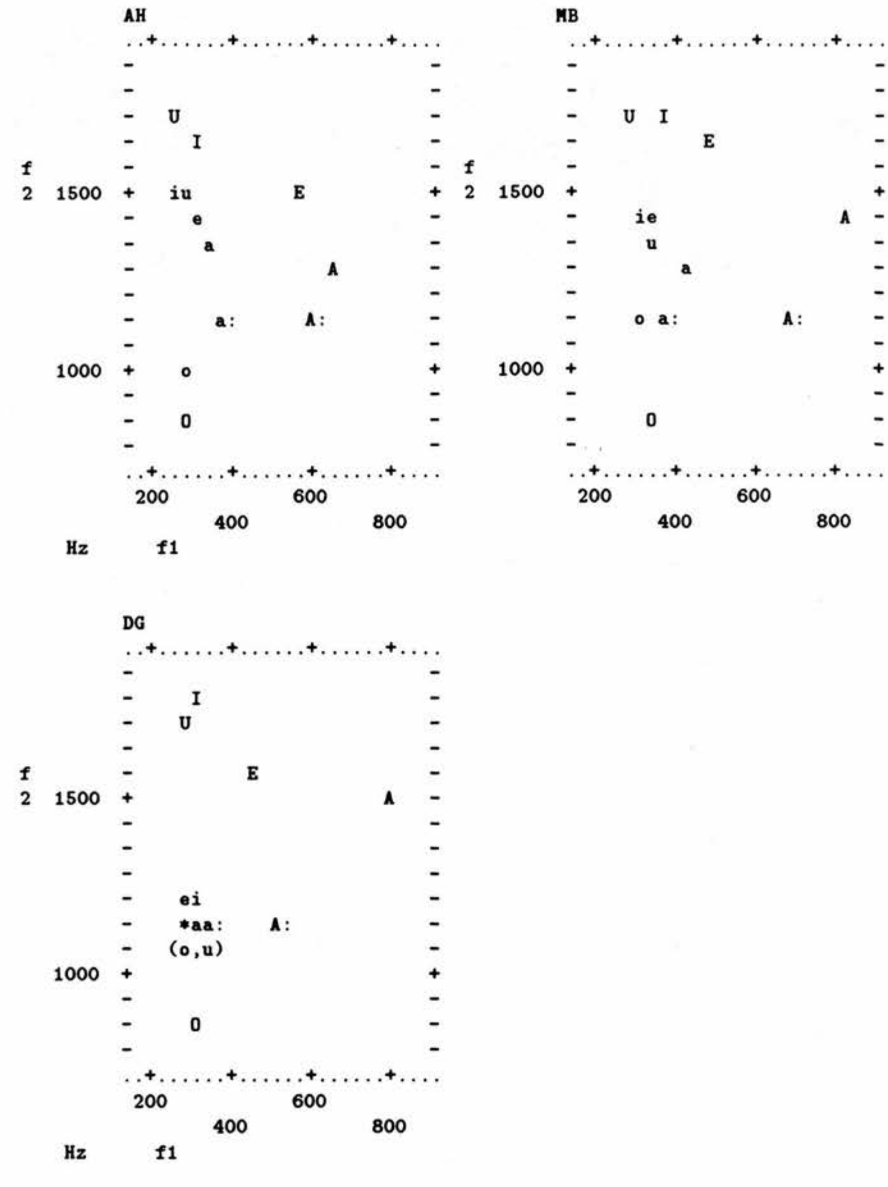



**Figure 3.12.** The scatterplots of the mean $F_1/F_2$ values of schwa in the contexts of /ɪ, ɛ, æ, ɑ, ɔ, u/ and the contextual vowels for the three speakers. Symbols i, e, a, a:, o, u = schwa in the contexts of /ɪ, ɛ, æ, ɑ, ɔ, u/. Symbols I, E, A, A:, O, U = contextual vowels /ɪ, ɛ, æ, ɑ, ɔ, u/. * means an overlap in value for the vowels in parentheses.

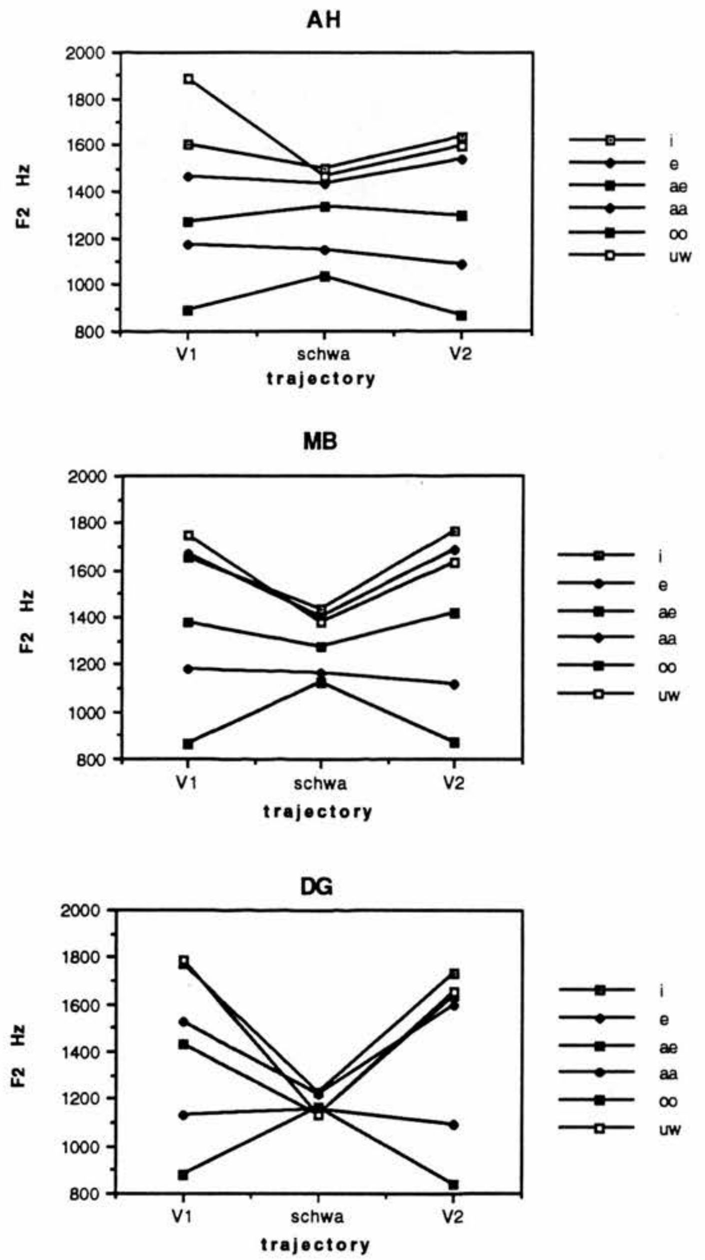



**Figure 3.13.** The mean $F_2$ trajectories from the preceding vowel to the following vowel across schwa in the VCəCV sequences for the three speakers. The consonantal context is /b/. Symbols i, e, ae, aa, oo, uw = contextual vowels /ɪ, ɛ, æ, ɑ, ɔ, u/.

speaker to speaker. Coarticulation may be idiosyncratic in fine detail.

**Weakening of Schwa**

In a number of cases, schwa was further weakened by shortening, devoicing or deletion. Figure 3.14 illustrates the cases of a fully voiced, partially voiced and completely deleted schwa. Dalby (1986) conducted an extensive study of the process of schwa deletion in connected speech. He observed that when syllable structures would have been affected by the deletion of schwa, significantly fewer cases of schwa deletion were observed. That is, schwa deletion is somehow constrained by the prosodic structure of the language, and schwa retains its [+syll, -cons] nature. Schwa devoicing or deletion seems to be a case of gestural overlap similar to vowel devoicing in Japanese. (See 5.5: p 106 for more discussion.) When schwa is devoiced or deleted, it seems to finally lose its vocalic identity, and it totally assimilates with the surrounding consonants.

### 3.2.3 Summary of results

The results of the experiment are summarized below.

1. The main effect of consonantal contexts was significant though small for the $F_1$ of schwa. No significant effects of vocalic contexts on $F_1$ were obtained. The mean $F_1$ values of schwa are 261 Hz, 321 Hz and 289 Hz for AH, MB and DG with the standard deviations of 39 Hz, 43 Hz, 38 Hz respectively. The standard deviations for schwa seem to be relatively small. However, from the observation of the $F_1$ trajectories in the present experiment, we may not conclude whether schwa is targeted or targetless in $F_1$. Though the $F_1$ trajectories do not stay level, they do not seem to be aiming at a common target, either.[12]

[12]The mean $F_1$ values at the midpoint of Dutch schwa for 3 speakers were 378(46) Hz, 315(37) Hz and 308(31) Hz for CəCV nonsense words and 358(50) Hz, 315(42) Hz and 347(45) Hz for VCəC nonsense words with the consonantal contexts of /p, t, k, f, s, χ, m, n, ɳ, r, l, j, ʋ/ and the vocalic contexts of /i, a:, u/ (Bergem 1994). The $F_1$ values observed for Dutch schwa are around 300 Hz and quite similar to the values observed for English schwa in the present study. (The values within the parentheses are standard deviations.)

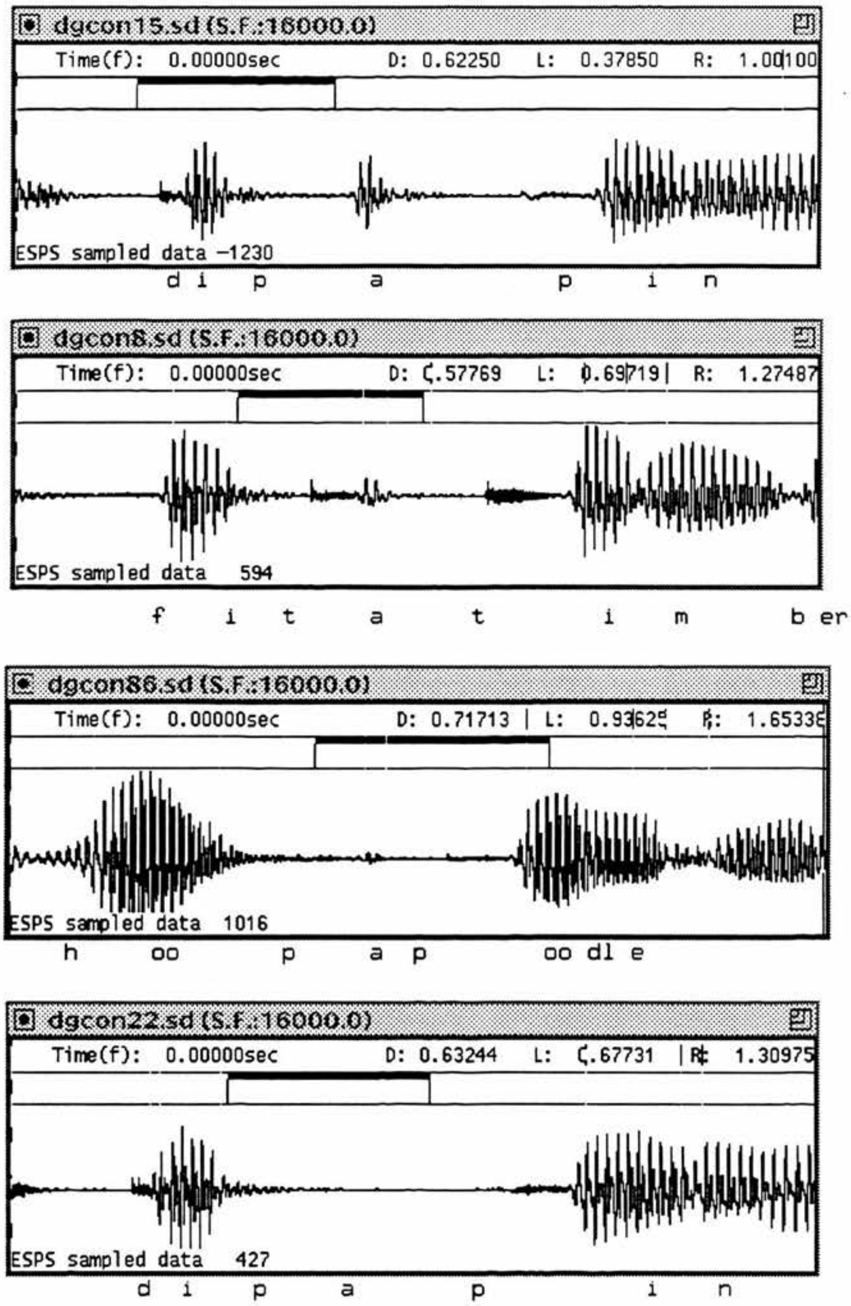



**Figure 3.14.** The four waveforms of schwa by DG. They illustrate examples of fully voiced, partially voiced and deleted schwa. The utterances are from the top: *dip a pin*, *fit a timber*, *hoop a poodle* and *dip a pin*.

2. The main effects of both consonantal and vocalic contexts were highly significant on the $F_2$ of schwa for the three speakers.

3. The effects of consonantal contexts were greater in magnitude than those of vocalic contexts. A large variability of schwa was observed not only at the transitions, but it was observed throughout the whole duration of schwa. The diversity of $F_2$ trajectories as a function of contexts and their levelness suggest that schwa may be targetless in $F_2$.

4. Interaction of consonantal and vocalic contexts was observed in the $F_2$ of schwa. The velar consonant /k/ assimilated with the contextual vowels, and schwa in the context of a fronted /k/ had a higher $F_2$ value, while schwa in the context of a back /k/ had a lower $F_2$ value. Different degrees of $F_2$ variability as a function of vocalic contexts were observed in different consonantal contexts. Some consonants seem to block V-to-V effects while others are more transparent to such effects.

5. In general, coarticulatory patterns observed on schwa were consistent across the three speakers. However, some idiosyncrasy in coarticulation was observed. For example, in the labial contexts of /p, b/ the three speakers showed different degrees of V-to-V coarticulation. Also, AH showed stronger V-to-V effects compared to the other speakers. On the other hand, he showed less C-to-V effects. Differences in the degree of coarticulation as well as preference for a certain type of coarticulation were observed across speakers.

The results as a whole suggest that schwa may be targetless in $F_2$. We could not come to a conclusion about the $F_1$ underspecification.

### 3.2.4 Discussion

The results of the experiment did not support the phonetic underspecification of schwa in $F_1$. No significant effects of the vocalic context were observed in $F_1$, and the effect of the consonantal context was small. Kuehn & Moll (1972) and Fowler (1981) also report little effects of vocalic contexts on the $F_1$ values of schwa. A perception test conducted by Kuehn & Moll (1972) showed that the schwa portion of the sequence /hædəCV/ contained 20 % information related to tongue advancement of the following vowel though the information relative to height was nearly nil. The greatest confusion in the identification of the following vowels was seen in the tongue height dimension. Fowler (1981) also reports that in the study of the effects of flanking vowels on the medial /ʌ/ in the sequences /VCʌVC/, both initial and final flanking vowels influenced the medial /ʌ/ in the expected directions on the second formant frequencies. However, on the first formant frequencies, no reliable effects on the medial vowel by the flanking vowels were observed.

The main effect of consonantal context on $F_1$ was statistically significant for all the subjects, but differences in the mean $F_1$ values as a function of different contextual consonants /p, t, k/ were rather small. Stevens and House (1963) report that shifts in average $F_1$ frequencies between 8 American English vowels in consonantal contexts and those in null environments (a /h_d/ context and in isolation) are fairly small. The general tendency was a lowering of $F_1$ values in consonantal contexts. As vowels are characterized by more open constriction, they may not reach the extreme low $F_1$ values of consonants. The $F_1$ values of schwa observed in the present experiment seem to reflect the default tongue height as a result of consonant release. In a number of cases where schwa was devoiced or deleted, e.g., *fit a timber* was pronounced as [fɪt$^h$t$^h$ɪmbə], there was always a release.

On the other hand, the $F_2$ values of schwa seem to be unspecified. The linear regression equation for the onset ($x$) and the midpoint ($y$) $F_2$ values of schwa is $y = 1.0011x - 13.05$ with the standard error of estimate of 69 Hz across three speakers. In this equation, the value of $y$ is very close to $x$. This suggests that the phonetic underspecification of schwa may be supported. The variability in $F_2$ observed at the vowel midpoint of schwa was great. Öhman (1966) reports that the range of mean $F_2$ values at the VC and CV transitions of the Swedish vowel /$\phi$/ in VCV utterances (both syllables were equally stressed) as a function of consonantal and vocalic contexts was around 600 Hz. On the other hand, he observed little influence of contexts at the steady state part of a vowel. At the stationary part of the vowel /$\phi$/ in Swedish, difference in mean $F_2$ values as a function of consonantal contexts /b, d, g/ was 70 Hz. In the present study, the contextual effects were observed right through the schwa. Even at the vowel midpoint, the range of mean $F_2$ values as a function of contexts remained about 700 Hz for schwa. This seems to illustrate the contrast between a specified and an unspecified vowel. Bates (forthcoming) has also shown that formant trajectories of full vowels diverge towards different contexts at the transitions, while at the midpoint, they cluster into a narrower band. On the other hand, a wide band of trajectories was observed throughout schwa. Her observation is based on a large collection of British English sentences from a single subject.

In general, $F_1$ is considered to be correlated with tongue height while $F_2$ is correlated with backness. In traditional phonological terms, then, it seems that schwa may be unspecified for [Backness]. The results of the experiment did not support the phonetic underspecification of schwa in $F_1$. However, the $F_1$ trajectories did not point to any target, either. On the other hand, the range of $F_1$ variability was small and low. Schwa seems to retain a minimally open tongue height in order to retain its vocalic identity. In this sense, schwa may be considered to be specified for [Height]. In other words, schwa may be specified as [+vocoid] by retaining the vocalic height while it is unspecified for the place of articulation.

There are cases where segments are specified only in [Height]. Kabardian, a Circassian language spoken in the Northeast Caucasus, has a vertical three-vowel system with three (high, mid, low) vowels /ɨ, ə, a:/ that are specified only for [Height]. They are redundantly distinguished by three-way durational contrast. That is, /ɨ/ is the shortest, while /a:/ is the longest and /ə/ comes in between the two in duration. In $F_2$, however, these vowels vary in quality as a function of their adjacent consonants to such an extent (the range of over 1000 Hz) that to a casual observer it would appear as if the language had a rich range of vowel qualities (Ladefoged & Maddieson 1990; Choi 1990). Choi presents a target-interpolation model of the $F_2$ trajectories of the Marshallese vowels using a least-squares curve-fitting procedure. The model captures the time-varying functions of the $F_2$ trajectories based on the effects of asynchronously timed primary and secondary consonantal constrictions, coupled with the effects of variation in constriction size associated with vowel height. The model captures the phonetic underspecification of the Marshallese vowels by not including a vocalic $F_2$ value. Another example of such partial specification is given by Keating (1988). She suggests that [s] may be specified for [Height] but unspecified for [Backness] (see page 16).

## 3.3 Conclusion

The nature of British English schwa was studied in the present chapter. It was concluded that acoustically, schwa is targetless in $F_2$. That is, schwa may be unspecified in [Backness]. The contextual effects on the $F_1$ values of schwa were small and there was small variability in $F_1$. However, the $F_1$ trajectories did not seem to converge to a common target value, either. On the other hand, schwa seems to have a certain vocalic height to retain its status of syllable nucleus. The $F_1$ values obtained in the present study were rather low and not central in value. This seems to suggest that schwa aims for a certain minimally open target height to retain its vocalic identity. In other words, schwa seems to be specified as [+syll, -cons] as suggested by Durand (1990), but unspecified in the place of articulation. The hypothesis proposed in the present experiment is thus partially supported. That is, schwa is a product of contextual assimilation in $F_2$. This supports the view that vowel reduction is contextual assimilation rather than centralization.

# Chapter 4

# Comparing Schwa and a Full Vowel

## 4.1 Variability of schwa and a full vowel

In the previous chapter, schwa was studied on its own. In the present chapter, the coarticulatory pattern of schwa will be compared with that of a full vowel. In Chapter 3, the partial underspecification of schwa was supported. It was suggested that schwa may be targetless in $F_2$. It was concluded that vowel reduction may be a result of contextual assimilation in $F_2$. That is, there is a contrast between a reduced vowel, a colourless vowel with no inherent $F_2$ target of its own and a full vowel which is targeted. In other words, there is a type of vowel that is extremely variable as a function of contexts on the one hand, while there is a type of vowel that is extremely resistant to contextual effects on the other hand. This contrast of targetedness in $F_2$ seems to be an important feature of so called stress-timed languages like English as will be discussed below. In order to show that this contrast exists in English, the following hypothesis is proposed.

- Hypothesis 1: The native speakers of English will manifest a sharp contrast in the extent of vowel variability in $F_2$ between full and reduced vowels as a function of contexts.

The variability of schwa and a full vowel /æ/ will be compared to test the above hypothesis. The hypothesis predicts that while schwa will be extremely variable, the full vowel /æ/ of English will be less variable.

In the present experiment only the second formant values will be studied as schwa was observed to be targetless only in $F_2$. The present experiment also limits itself to the study of V-to-V effects. It would be desirable to compare both C-to-V and V-to-V effects on schwa and a full vowel with the same controlled sequence of the VC_CV context used in Chapter 3. However, this had to be given up because of the problem of placing a full vowel into the VC_CV paradigm in natural sentences. This is due to the nature of the English rhythm itself. Instead, Vb_bV sequences were used with /ə/ and the full vowel /æ/ as the middle vowel. For the contextual vowels, the vowels /ɪ/ and /æ/ were used. For the VCæCV sequences, the reduced vowels /ɪ/ and /ə/ were used for the following contextual vowel. Both symmetric and asymmetric contexts were considered to observe the difference in the relative strength of carryover and anticipatory coarticulation.

## 4.2 Transparency of schwa and stress timing

While limiting the study to V-to-V effects seems to be a disadvantage when strong consonantal effects have been observed on schwa, the difference in the extent of V-to-V effects may be a good measure to illustrate the contrast of a reduced vowel and a full vowel. In the spectrographic study of VCV articulation in Swedish and English, Öhman (1966) observed the effects of the first vowel at the onset of the second vowel and the effects of the second vowel at the offset of the first vowel.[1] The $F_2$ trajectories across the medial consonant looked continuous with the consonant portion superimposed on it. From this observation, he described the VCV articulation as the basic diphthongal gesture of vowels with the consonantal gesture superimposed on the transitional portion. This seems to imply that speech is organized from vowel to vowel. That is, there is a basic rhythm which spans from vowel to vowel, on top of which other segments are superimposed or coproduced.

[1]In Öhman's experiment, both vowels were equally stressed.

Smith (1991) studied the timing of the articulatory movements of VCV utterances with a single and a geminate C between the vowels for a so called syllable-timed (Italian) and mora-timed (Japanese) language (an X-ray microbeam study). The results of her study suggest that the timing of the two vowels relative to each other is controlled independently of the consonant in Italian. That is, the timing of the vowel to vowel articulation did not change whether there was a single consonant or a geminate between the two vowels. On the other hand, in Japanese, the articulatory movements for the second vowel were delayed when the medial consonant was a geminate. Her results imply that in so called syllable timed languages the basic rhythm spans from vowel to vowel or from syllable nucleus to syllable nucleus.

The above results lead us to the question: what is the rhythm of so called stress-timed languages like? Fowler (1981) developed the idea presented by Öhman (1966) above to stressed vowels. That is, the basic rhythm of English may span from stressed vowel to stressed vowel. All the other segments are superimposed on top of this basic diphthongal articulation. Fowler suggests that

> ...talkers organize the production of relatively long stretches of speech - that is, the intervals between stressed syllables within a phrase - more-or-less as they organize the production of syllables. Their strategy perhaps, is to subsume the production of segments other than stressed vowels within the domain of production of a stressed vowel. Indeed, this subsumption strategy may account for the linguists' impression that English is stress-timed (Fowler 1981:p128).

As the effects of the first vowel were observed at the onset of the second vowel and vice versa in the VCV articulation, Fowler's hypothesis would predict that the effects of the first stressed vowel may be observed at the onset of the second stressed vowel and vice versa in the VCəCV utterances. That is, the contextual effects may pass right through the schwa and the two consonants from the preceding stressed vowel to the following stressed vowel. This may also happen in the other direction. Such long-term effects have been reported by Huffman (1986) and Magen (1989). Magen observed such effects in /bVbəbVb/ utterances. Huffman observed the effects of a first vowel on a third across two /l/'s and a /ə/ though she did not observe such long-term effects when the intervening consonants were /d/'s.[2]

The transparency of schwa observed in Experiment 1 supports Fowler's hypothesis. In the VCV articulation, the C may be transparent to the V-to-V effects though the transparency of consonant to vocalic effects seems to vary from consonant to consonant. (See page 59. See also page 23 for more discussion.) In the symmetric VCəCV utterances studied in Chapter 3, the CəC portion as a whole may be transparent to the V-to-V effects. Indeed the $F_2$ trajectories of schwa observed in Figure 3.6 suggest that the CəC as a whole seems to be acting as if it was a single unit or even a long consonant.

As the V-to-V effects may penetrate through the intervening consonant in the VCV articulation, the V-to-V effects may penetrate through the medial schwa in the VCəCV articulation. Therefore, the nature of V-to-V effects observed on schwa and a full vowel will be essentially different. The V-to-V effects will be observed right through the schwa, while the V-to-V effects may not penetrate through a stressed vowel. This is because the interval from stressed vowel to stressed vowel may be a basic production unit of English in which the onset of the second stressed vowel may belong to both the preceding and the current unit. The idea is similar to the ambisyllabicity. The VCV articulation where the two V's are full vowels may be considered as a subset of the $VC_n$(unstressed V)($C_n$)V articulation.

Magen (1984) observed the effects of the transconsonantal vowel throughout the unstressed full vowel in the VbV utterances where either the initial or the final vowel was stressed and the other was unstressed. The vowels were /i/ or /a/, and they were not reduced. Thus, the transparency of a vowel may be a common feature of unstressed vowels whether it is reduced or not though the degree of transparency may be greater for schwa than unstressed full vowels. For stressed vowels, the V-to-V effects diminished towards the middle of the vowel.[3] From these observations, the second set of hypotheses may be proposed.

[2]That the alveolar obstruent /d/ blocked the long-term effects is consistent with the results of the present study. In Experiment 1, the alveolar obstruent /t/ was observed to block V-to-V effects.

[3]Magen did not perform any statistical analyses to test the significance of the difference between the $F_2$ trajectories in the /i/ and /a/ contexts.

- Hypothesis 2a: The V-to-V effects will be observed right through the schwa from the onset to the offset and vice versa in both directions.
- Hypothesis 2b: The V-to-V effect will not penetrate through a stressed vowel in either direction.

# 4.3 Experiment 2

## 4.3.1 Methods

### Material

The following two sets of sentences were used for (a) VCəCV and (b) VCæCV sequences. Each subject repeated the sentences 5 times in a randomized order.

(a) VCəCV

1. ɪbəbɪ: The campaign for Women's **Lib** **aby**smally failed.
2. ɪbəbæ: We found the baby's c**rib** **aba**ndoned in the car park.
3. æbəbɪ: The inspector considered the l**ab** **aby**smal.
4. æbəbæ: The cr**ab** **aba**ndoned its prey as it sensed something approaching.

(b) VCæCV

1. ɪbæbɪ: The **fib** **Abbey** National's TV advert was said to contain turned out to be quite legal.
2. ɪbæbə: When today's students were in the c**rib** **Abba** were superstars.
3. æbæbɪ: The robbers planned to gr**ab** **Abbey** National's armoured van.
4. æbæbə: Nostalgia fans like to gr**ab** **Abba** records when they see them.

| | ə | | | æ | | |
|---|---|---|---|---|---|---|
| | onset | midpoint | offset | onset | midpoint | offset |
| pre_v (p) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0032 | 0.3924 |
| fol_v (f) | 0.0003 | 0.0001 | 0.0000 | 0.1670 | 0.9541 | 0.0000 |
| speaker (s) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| p × f | 0.0323 | 0.0470 | 0.4792 | 0.7323 | 0.5361 | 0.1543 |
| p × s | 0.0000 | 0.0238 | 0.2487 | 0.0558 | 0.1862 | 0.4104 |
| f × s | 0.3274 | 0.1792 | 0.0265 | 0.5731 | 0.0002 | 0.7602 |
| p × f × s | 0.6466 | 0.0664 | 0.4564 | 0.7846 | 0.3936 | 0.2321 |

**Table 4.1.** The level of significance of the results of ANOVAs. The table shows the main effects of the preceding vowel, following vowel and speaker, and their interaction on the second formant values of schwa and the vowel /æ/ across five speakers.

**Speaker**

Eight male native speakers of British English participated in the present experiment. They were AH, MB, DN, JP, JD, GW, DP and RL. AH and MB participated in Experiment 1 as well. AH is originally from Manchester. MB is from Lincolnshire. DN is originally from Liverpool. JP is from Kent. JD is from Buckinghamshire. GW is from Hastings. DP is from London and RL is from New Castle. AH speaks standard North Western accent. The other speakers speak RPish standard Southern British accent. At the time of the recording they were either undergraduate or postgraduate students at the Department of Linguistics, University of Edinburgh. They age between 20 to 31.

**Recording and Analysis** Recording and Analyses were conducted in the same procedure as described above in Experiment 1. (See 3.2.1: p 42.)

### 4.3.2 Results

Three-way ANOVAs were performed with the $F_2$ values of schwa and the vowel /æ/ at the onset, midpoint and offset of the segment as dependent variables. The independent variables were the preceding vowel, following vowel and speaker. Statistical analyses were performed with the pooled data across speakers. Table

4.1 summarizes the results.

For schwa, the main effects of the preceding vowel, following vowel and speaker are significant at the three points in the trajectory. Significant interactions were observed between the preceding and following vowel at the onset and midpoint. Significant interactions were also observed between the preceding vowel and speaker at the onset and midpoint and between the following vowel and speaker at the offset (Appendix A). The speakers showed different magnitudes of V-to-V effects in these cases. For MB, the mean $F_2$ value in the context of the following /æ/ was higher than that in the context of /ɪ/ at the offset.

For the full vowel /æ/, the main effect of the preceding vowel was significant at the onset and the midpoint, and the main effect of the following vowel was significant only at the offset. The main effect of the speaker was significant throughout the trajectory. Significant interaction was observed only between following vowel and speaker at the midpoint (Appendix A). Where the interaction was significant, speakers showed different magnitudes of V-to-V effects. JD and DP showed a reverse trend from the other speakers as a function of the following vowel at the midpoint; i.e., their mean $F_2$ values were higher in the context of /æ/ than /ɪ/.

The results above show the transparency of schwa compared to that of the full vowel /æ/. The effects of the contextual vowels were observed right through the schwa in both directions. On the other hand, for /æ/, the effect of the preceding vowel was observed only at the onset and midpoint, and the effect of the following vowel was observed only at the offset of the vowel /æ/. The results also seem to suggest that the strength of the V-to-V effects is greater for carryover than for anticipatory effects. While the effects of the following vowel seem to be stopped at the offset, the effects of the preceding vowel seem to penetrate into the midpoint of the full vowel /æ/. The stress pulse seems to be strong enough to be carried onto the midpoint of the stressed syllable of the next stress foot.

Figure 4.1 shows the coefficient of variation for the $F_2$ values of schwa and the vowel /æ/ at the vowel midpoint for each of the eight speakers. The coefficient of variation is the ratio of the standard deviation and the mean. Thus, higher coefficient of variation suggests large variability. For most speakers, the coefficient of variation is higher for schwa than for the vowel /æ/. However, two speakers,

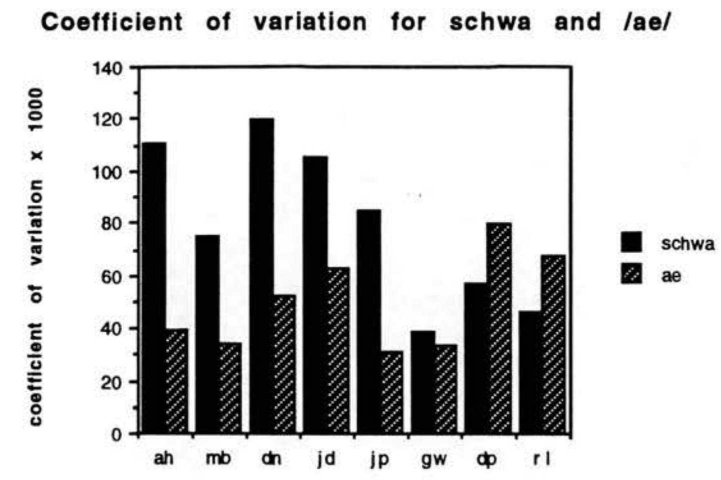



**Figure 4.1.** The coefficient of variation of the $F_2$ values of schwa and the vowel /æ/ at the vowel midpoint for each speaker.

DP and RL, show the opposite trend. The speaker GW also shows very little difference. For RL, the mean $F_2$ values of schwa was 1431.2 Hz in the symmetric /ɪ/ context and 1295.9 Hz in the /æ/ context. For his vowel /æ/, the mean $F_2$ values were 1461.2 Hz in the symmetric /ɪ/ context and 1334.6 Hz in the /æ_ə/ context. The differences in the mean $F_2$ values as a function of vocalic contexts are 135.3 Hz for schwa and 126.6 Hz for /æ/. Thus, his schwa still shows slightly greater effects of the contextual vowels. For DP, the mean $F_2$ values of schwa was 1326.9 Hz in the /ɪ/ context and 1212.3 Hz in the /æ/ context. The difference is 114.6 Hz. On the other hand, he had higher mean $F_2$ value in the context of /æ_ə/ (1414.3 Hz) than in the context of /ɪ/ (1377.3 Hz) for the full vowel /æ/. Thus, these two speakers still seem to show greater systematic variability for schwa than for the full vowel /æ/.

Figures 4.2 through 4.6 illustrate the difference in the coarticulatory pattern of /ə/ and the vowel /æ/.[4] The effects of the vocalic contexts are greater on schwa than on the full vowel /æ/. Figure 4.2 shows the $F_2$ trajectories across eight speakers in the symmetric /ɪb_bɪ/ and /æb_bæ/ (/æb_bə/) contexts for the vowels /ə/ and /æ/. Different magnitudes of contextual effects are clearly observable. The trajectories for /æ/ seem to merge at the vowel midpoint for the target value.

Figures 4.3 and 4.4 show carryover V-to-V effects through /ə/ and /æ/. The effects of the preceding vowel are observed right through the schwa, while such effects are observed at the onset and somewhat marginally at the midpoint for /æ/. The $F_2$ trajectories in Figure 4.4(b) also diverge towards the offset during the following schwa. This is because of the effects of the consonants /r/ and /w/ that come after the schwa (*Abba records* vs. *Abba were*). The /w/ is pulling down the $F_2$ value. Though not significant, the effects of these diverse schwa trajectories are observed at the offset of the medial vowel /æ/.

Figures 4.5 and 4.6 show anticipatory V-to-V effects. The effects of the following vowel are observed right through the schwa, while such effects are observed only at the offset for /æ/.

[4]Each data point in these figures is the mean of 5 repetitions × 8 speakers. In some cases there are less than 40 observations per mean due to formant tracking error. The smallest number of observations per mean is 34.

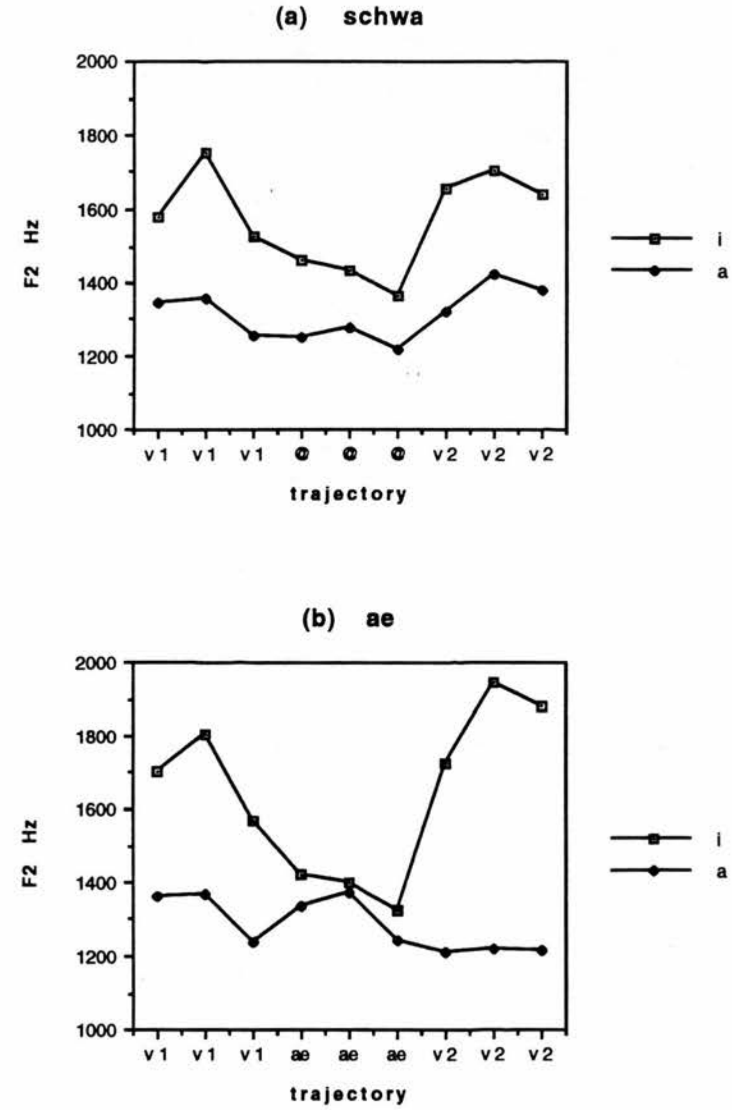



**Figure 4.2.** The mean $F_2$ trajectories across 8 speakers of (a) the VbəbV and (b) VbæbV sequences in the symmetric contexts of /ɪb_bɪ/ and /æb_bæ/ (/æb_bə/).

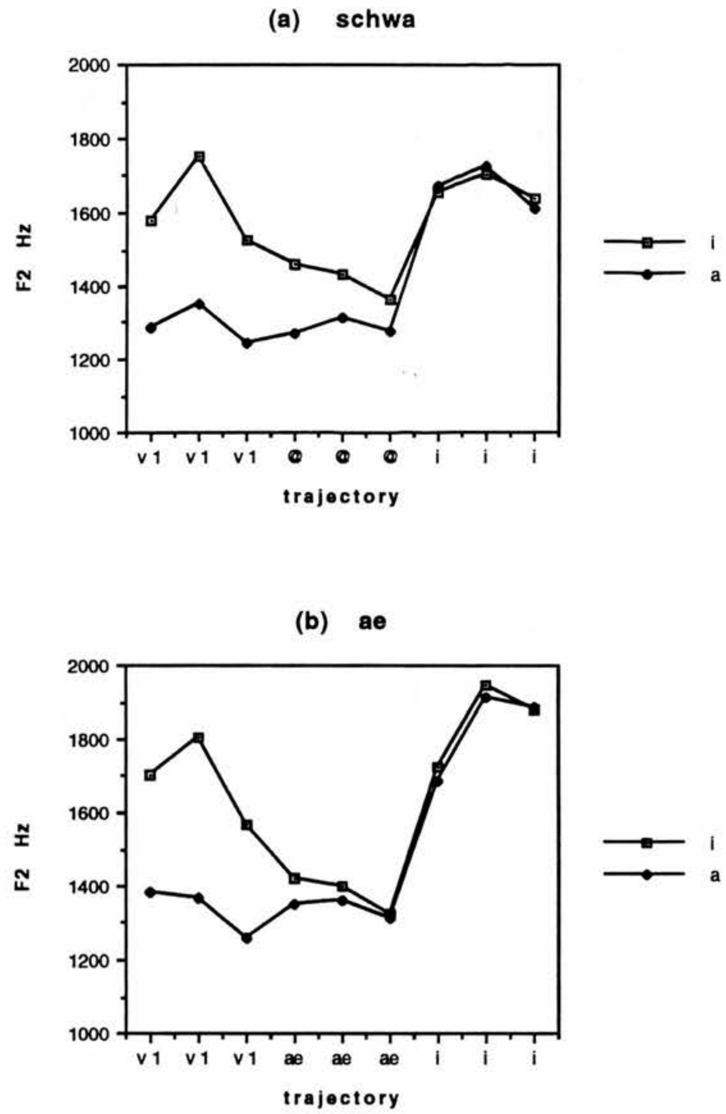



**Figure 4.3.** The mean $F_2$ trajectories across 8 speakers of (a) the VbəbV and (b) VbæbV sequences in the asymmetric contexts of /Vb_bɪ/. The effects of the preceding stressed vowels /ɪ/ and /æ/ are observed.

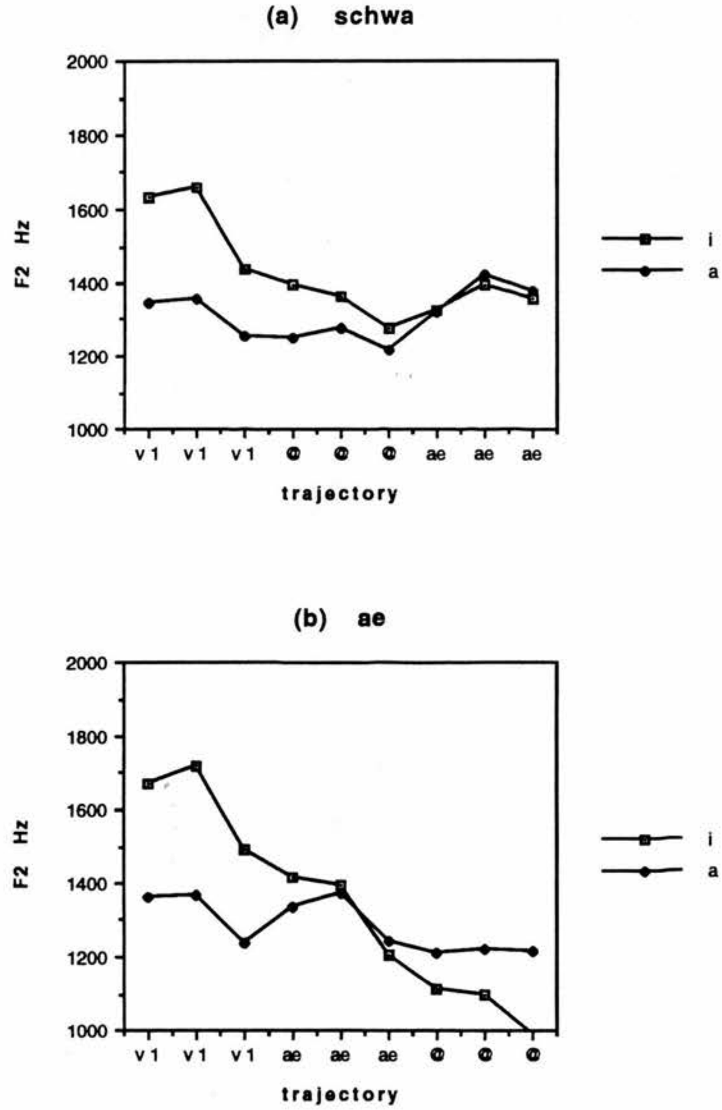



**Figure 4.4.** The mean $F_2$ trajectories across 8 speakers of (a) the VbəbV and (b) VbæbV sequences in the asymmetric contexts of /Vb_bæ/ (/Vb_bə/). The effects of the preceding stressed vowels /ɪ/ and /æ/ are observed.

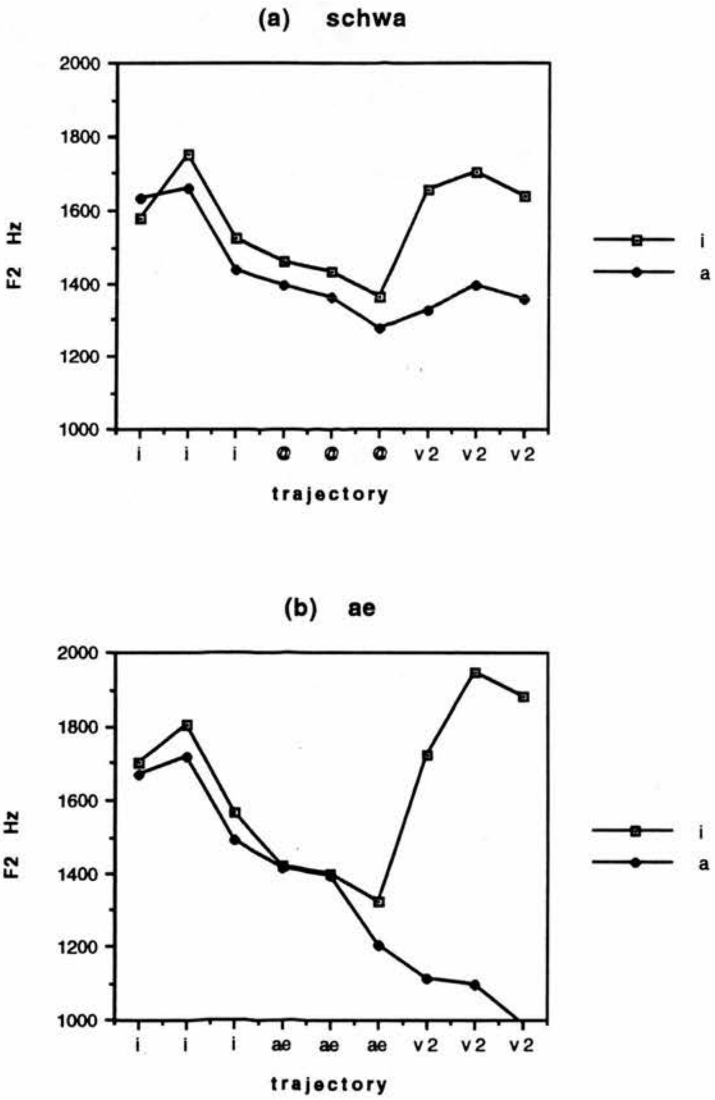



**Figure 4.5.** The mean $F_2$ trajectories across 8 speakers of (a) the VbəbV and (b) VbæbV sequences in the asymmetric contexts of /ɪb_bV/. The effects of the following vowels /ɪ/ and /æ/ (or /ə/), stressed or unstressed, are observed.

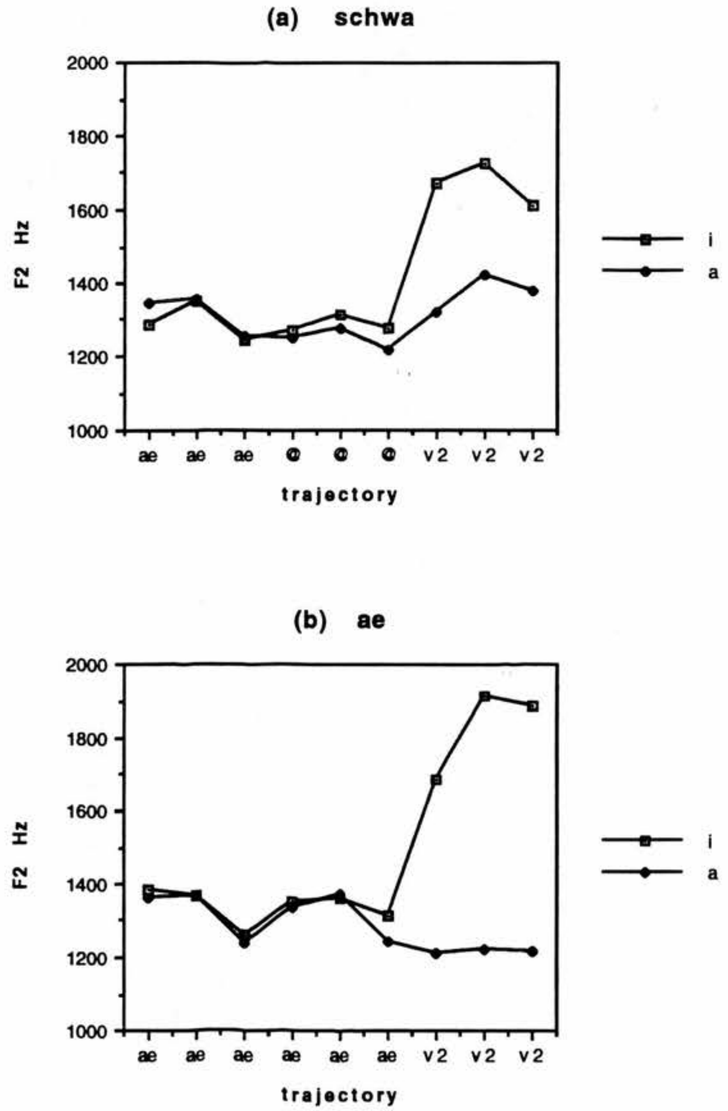



**Figure 4.6.** The mean $F_2$ trajectories across 8 speakers of (a) the VbəbV and (b) VbæbV sequences in the asymmetric contexts of /æb_bV/. The effects of the following vowels /ɪ/ and /æ/ (or /ə/), stressed or unstressed, are observed.

| Symmetric: Vb_bV | | | | |
|---|---|---|---|---|
| Vowel type | contexts | onset | midpoint | offset |
| ə | ɪ | 1461.6 | 1434.9 | 1360.2 |
| | æ | 1247.3 | 1276.7 | 1214.4 |
| | ɪ− æ | *214.3 | *158.2 | *145.8 |
| æ | ɪ | 1424.0 | 1401.8 | 1322.2 |
| | æ(ə) | 1333.3 | 1370.4 | 1244.1 |
| | ɪ− æ(ə) | 90.7 | 31.4 | 78.1 |

**Table 4.2.** The mean $F_2$ values across 8 speakers of the vowels /ə/ and /æ/ at the onset, midpoint and offset of the segment in the symmetric /ɪb_bɪ/ and /æb_bæ/ (/æb_bə/) contexts. The differences in $F_2$ values as a function of vocalic contexts are also shown. The symbol * indicates that the difference has reached statistically significant level ($p < 0.05$).

When Figures 4.3 & 4.4 and Figures 4.5 & 4.6 are compared, it becomes clear that carryover effects are greater than anticipatory effects, particularly for schwa.

Tables 4.2 through 4.6 show the $F_2$ values that were inputs to Figures 4.2 through 4.6. For example, Table 4.2 shows the $F_2$ values of schwa and the full vowel /æ/ in the symmetric (or pseudo-symmetric) contexts of /ɪb_bɪ/ and /æb_bæ/ or /æb_bə/ at the onset, midpont and offset of the segment. The table also shows the difference in $F_2$ values as a function of the contextual vowels /ɪ/ and /æ/ (or /ə/). The symbol * indicates that the difference has reached a statistically significant level ($p < 0.05$).[5] Table 4.3 shows the $F_2$ values of /ə/ and /æ/ when the third vowel was held constant to /ɪ/. The effects of the preceding vowel at the onset, midpoint and offset of the vowels /ə/ and /æ/ may

[5] Post hoc scheffe tests were performed for the interaction between the contextual vowels (/I/ vs. /æ/ or /ə/), the points of measurement (onset, midpoint or offset) and the affected vowels (/ə/ or /æ/) for the five subsets of the data. The five subsets are one symmetrical context /Ib_bI/ vs. /æb_bæ/ (/æb_bə/) and four asymmetrical contexts /Ib_bI/ vs. /æb_bI/, /Ib_bæ/ vs. /æb_bæ/ (/Ib_bə/ vs. /æb_bə/), /Ib_bI/ vs. /Ib_bæ/ (/Ib_bə/) and /æb_bI/ vs. /æb_bæ/ (/æb_bə/). There are 2 (contextual vowels) × 3 (points of measurement) × 2 (affected vowels) × 8 (speakers) × 5 (repetitions) = 480 tokens per subset. There are 12 cell means and 40 observations per mean. In some cases there were less than 40 observations per mean due to formant tracking error. The smallest number of observations per mean was 36.

| Carryover: Vb_bɪ | | | | |
|---|---|---|---|---|
| Vowel type | preceding vowel | onset | midpoint | offset |
| ə | ɪ | 1461.6 | 1434.9 | 1360.2 |
| | æ | 1270.5 | 1316.1 | 1277.3 |
| | ɪ− æ | *191.1 | *118.8 | 82.9 |
| æ | ɪ | 1424.0 | 1401.8 | 1322.2 |
| | æ | 1350.3 | 1362.7 | 1312.4 |
| | ɪ− æ | 73.7 | 39.1 | 9.8 |

**Table 4.3.** The mean $F_2$ values across 8 speakers of the vowels /ə/ and /æ/ at the onset, midpoint and offset of the segment in the asymmetric /ɪb_bæ/ (/ɪb_bə/) and /æb_bæ/ (/æb_bə/) contexts. The differences in $F_2$ values as a function of the preceding vowel are also shown. The symbol * indicates that the difference has reached statistically significant level ($p < 0.05$).

| Carryover: Vb_bæ(ə) | | | | |
|---|---|---|---|---|
| Vowel type | preceding vowel | onset | midpoint | offset |
| ə | ɪ | 1393.0 | 1360.8 | 1278.0 |
| | æ | 1247.3 | 1276.7 | 1214.4 |
| | ɪ− æ | *145.7 | 84.1 | 63.6 |
| æ | ɪ | 1414.9 | 1394.2 | 1205.0 |
| | æ | 1333.3 | 1370.4 | 1244.1 |
| | ɪ− æ | 81.6 | 23.8 | -39.1 |

**Table 4.4.** The mean $F_2$ values across 8 speakers of the vowels /ə/ and /æ/ at the onset, midpoint and offset of the segment in the asymmetric /ɪb_bæ/ (/ɪb_bə/) and /æb_bæ/ (/æb_bə/) contexts. The differences in $F_2$ values as a function of the preceding vowel are also shown. The symbol * indicates that the difference has reached statistically significant level ($p < 0.05$).

| Anticipatory: ɪb_bV | | | | |
|---|---|---|---|---|
| Vowel type | following vowel | onset | midpoint | offset |
| ə | ɪ | 1461.6 | 1434.9 | 1360.2 |
| | æ | 1393.0 | 1360.8 | 1278.0 |
| | ɪ− æ | 68.6 | 74.1 | 82.2 |
| æ | ɪ | 1424.0 | 1401.8 | 1322.2 |
| | ə | 1414.9 | 1394.2 | 1205.0 |
| | ɪ− ə | 9.1 | 7.6 | *117.2 |

**Table 4.5.** The mean $F_2$ values across 8 speakers of the vowels /ə/ and /æ/ at the onset, midpoint and offset of the segment in the asymmetric /ɪb_bɪ/ and /ɪb_bæ/ (/ɪb_bə/) contexts. The differences in $F_2$ values as a function of the following vowel are also shown. The symbol * indicates that the difference has reached statistically significant level ($p < 0.05$).

| Anticipatory: æb_bV | | | | |
|---|---|---|---|---|
| Vowel type | following vowel | onset | midpoint | offset |
| ə | ɪ | 1270.5 | 1316.1 | 1277.3 |
| | æ | 1247.3 | 1276.7 | 1214.4 |
| | ɪ− æ | 23.2 | 39.4 | 62.9 |
| æ | ɪ | 1350.3 | 1362.7 | 1312.4 |
| | ə | 1333.3 | 1370.4 | 1244.1 |
| | ɪ− ə | 17.0 | -7.7 | 68.3 |

**Table 4.6.** The mean $F_2$ values across 8 speakers of the vowels /ə/ and /æ/ at the onset, midpoint and offset of the segment in the asymmetric /æb_bɪ/ and /æb_bæ/ (/æb_bə/) contexts. The differences in $F_2$ values as a function of the following vowel are also shown. The symbol * indicates that the difference has reached statistically significant level ($p < 0.05$).

| Vowel type | AH | MB | DN | JD | JP | GW | DP | RL |
|---|---|---|---|---|---|---|---|---|
| ə | 51 | 57 | 43 | 41 | 37 | 46 | 35 | 41 |
| æ | 74 | 65 | 61 | 83 | 69 | 67 | 84 | 91 |

**Table 4.7.** The mean durations in ms of the schwa and the vowel /æ/ for the eight subjects.

be observed by looking at this table.

These tables show that though the contextual effects seem to penetrate through schwa in Figures 4.2 through 4.6, the differences in $F_2$ values as a function of vocalic contexts are not always statistically significant. This may be due to the small number of tokens used in the present experiment. For example, the L-to-R effects of the preceding vowel do not even reach the midpoint of schwa when the third vowel is /æ/. The R-to-L effects of the following vowel was not significant at all for schwa. This seems to suggest stronger carryover than anticipatory effects for schwa. For the full vowel /æ/, the difference in $F_2$ values as a function of vocalic contexts was significant only at the vowel offset for the anticipatory effects when the first vowel was /ɪ/. Rather strong effects of the following vowel observed at the offset of the vowel /æ/ compared to /ə/ in Tables 4.5 and 4.6 may be due to different $F_2$ values of the contextual vowels. The mean $F_2$ values of the following /ɪ/ and /æ/ for schwa were 1712.6 Hz and 1403 Hz, while the mean $F_2$ values of the following /ɪ/ and /ə/ for /æ/ were 1937.8 Hz and 1151.2 Hz respectively. That is, the two contextual vowels had a wider range of $F_2$ values for /æ/ than for /ə/.

In the present experiment, no significant effects of the first vowel were observed at the onset of the second vowel and vice versa across the /bəb/ segments.[6]

Another factor that may be crucially relevant to the issue of transparency is segment duration. As schwa is relatively short compared to full vowels, the effects of the transconsonantal vowels may penetrate through them more readily

[6]The long term V-to-V effects were observed for the Japanese vowel /a/. The effect of the third vowel was observed at the midpoint of the first vowel /a/ across two /b/'s and an /a/. Similarly, the effect of the first vowel was observed at the onset of the third vowel across the /bab/ sequence. See Figures 7.6 and 7.7.

than through full vowels. Table 4.7 shows the mean durations of the schwa and the full vowel /æ/ for each speaker. These mean durations suggest that the offset of schwa may correspond in time to the midpoint of the full vowel /æ/ for some speakers like JD, JP, DP and RL.

On the other hand, the segmental durations of the Japanese vowel /a/ and the English full vowel /æ/ are similar (see Table 7.9), and yet the two vowels exhibit quite different patterns of coarticulation (See Chapter 7). The Japanese vowel /a/ shows greater extent of V-to-V effects both in magnitude and in temporal extent compared to the English full vowel /æ/. This seems to suggest that while at the lower level of speech production, the time element may affect the extent of contextual effects, the general pattern of coarticulation may be determined at the higher level of the linguistic representation. The generally shorter duration of schwa, however, may be an important feature in order to guarantee the transparency of schwa.

### 4.3.3 Summary of results

1. The effects of the vocalic context were generally stronger on /ə/ than on the full vowel /æ/.

2. In general, the V-to-V effects were observed throughout schwa in both directions.

3. The effects of the preceding vowel were observed only at the onset and midpoint of the vowel /æ/ and the effects of the following vowel were observed only at the offset of the vowel /æ/.

4. For both schwa and /æ/, carryover V-to-V effects seem to be stronger than anticipatory effects.

These results support Hypotheses 1 and 2a & b proposed above.

### 4.3.4 Discussion

The results of the present experiment suggest that there is clear difference in the behaviour of schwa and the full vowel /æ/ in the conductivity of V-to-V effects.

The V-to-V effects penetrate right through schwa, while such effects are observed at the onset and midpoint for carryover effects and at the offset for anticipatory effects for the full vowel /æ/. This seems to support Fowler's hypothesis. That is, in English the basic speech production unit is an interval from stressed vowel to stressed vowel. As the effects of the first vowel may reach the onset of the second vowel across a single consonant in the VCV articulation, the effects of the first stressed vowel may extend as far as the onset of the second stressed vowel across a number of segments in $\acute{V}C_n\text{ə}C_n\acute{V}$ utterances.

If the rhythm of English is determined by the basic beat from stressed to stressed vowel, it seems intuitively correct that carryover effects are stronger. The effects of a stressed vowel seem to decay across time towards the end of the unit and a new pulse starts at the beginning of the next stressed vowel. That the effects of the first vowel are observed at the onset of the second vowel across a consonant in the VCV articulation suggests that the onset of a vowel may be "ambisyllabic" in nature. That is, they belong to both the preceding and current production unit. In the present experiment, the effects of the first stressed vowel were not observed at the onset of the second stressed vowel across the /bəb/ segments. However, by using more transparent consonants such as /h/ or /k/, such effects may have been observed. When there was only a single consonant between two stressed vowels, the carryover V-to-V effects could be observed to extend to the midpoint of the second stressed vowel. Also, the difference in the extent of V-to-V effects between /æ/ and /ə/ was greater for carryover effects than for anticipatory effects.

According to Fowler's hypothesis, segments between the stressed vowels are coproduced with the basic diphthongal articulation of the stressed vowels, and thus are transparent to the V-to-V gesture. In other words, as long as the articulation of the stressed vowels and the intervening segments are compatible with one another, they are coarticulated. Schwa plays an important function in this system. That schwa is absolutely transparent is an essential part of the speech production strategy of English and possibly other stress-timed languages. The transparency of schwa guarantees the coproduction of segments across a longer stretch of time. Compared to syllable-timed languages in which only a few consonants may intervene between the vowels, in stress-timed languages, a greater number of segments may intervene between the stressed vowels. For example, in the utterance *like to grab* /laɪktəgræb/, there are five segments between the full vowels /aɪ/ and /æ/. Syllable-timed languages and stress-timed languages may employ a similar organizational strategy. In both cases, there is essentially a vowel-to-vowel production unit. However, in stress-timed languages, vocalic elements as well as consonants may come between the vowels, and importantly these vowels are unspecified in $F_2$ and therefore are transparent to coarticulatory effects.

It may also be important that schwa retains its vocalic identity so that it could still be a syllable nucleus because the existence of these weak vowels seems to essentially differentiate so called stress-timed languages from syllable-timed languages. In the present study only schwa was studied. However, in the future, it would be necessary to look into the other reduced vowel in English, namely /ɪ/. The existence of syllabic consonants may also be an important feature of stress-timed languages. The transparency of unstressed full vowels must also be studied.

Another important issue is how the temporal effect interacts with the transparency of schwa. Vowel reduction correlates with duration (Lindblom 1963). Lindblom's (1963) model of vowel reduction predicts detailed variation in Swedish vowels given information about their durations and adjacent consonant loci. Engstrand (1992) also showed that the vowel duration and the difference in $F_2$ maximum and offset are highly correlated (r = 0.96) for natural discourse data. That is, as the duration gets shorter, the $F_2$ value of a vowel gets less targeted.

However, it seems that certain phonological and pragmatic rules may override this effect. Gay (1978) showed that stress and duration are two independent variables of vowel reduction that are not directly correlated with one another. He observed the extent of vowel reduction at two speech rates, fast and slow, for stressed and unstressed (not reduced) vowels.

> The slow and fast pairs within each stress condition are characterized by essentially the same overall amplitude, fundamental frequency and

> first and second formant frequencies. However, the corresponding unstressed syllables at each rate are consistently lower in overall amplitude and fundamental frequency and somewhat reduced in vowel colour. (Gay 1978:p228)

Thus, in circumstances where a stressed vowel gets shortened, it may still retain its prominence. Engstrand (1992) also discusses a case where his subject invariably reduced the vowel /i:/ in the adverbial phrase *på på vis* (in any way) irrespective of duration. This seems to suggest that there is a higher level pragmatic constraint. Moon (1991) also showed duration-dependent undershoot for both citation-form speech (at a comfortable rate and vocal effort) and clear (over-articulated) speech. Moon obtained a compact description of the relationship between the formant and duration by fitting an exponential curve to the formant versus duration plots. However, he had to adjust the values for the constants and the $F_2$ target value for the two different speech styles. This seems to suggest that within each speech style there is a systematic relationship between formant undershoot and duration. However, these relationships seem to differ from one speech style to another.

Lindblom (1990) describes the general reduction processes by the term hyper- versus hypo- articulated speech (H & H). A number of different terminology are used to express the same concept such as reduction vs. elaboration, casual vs. formal or careful vs. sloppy. The independence of the H & H dimension and the duration has been pointed out by a number of researchers (Zwicky 1972; Shockey 1973; Lindblom 1983; Engstrand 1988; Lindblom & Moon 1988; Moon & Lindblom 1989). That is, slow and reduced as well as fast and elaborated speech have been described in literature. For example, Engstrand (1988) showed that short vowel durations elicited by a fast speech rate do not necessarily lead to formant undershoot. Nord (1986) also showed that long duration due to final lengthening can be compatible with considerable reduction.

Engstrand also suggests a possibility that shorter duration may be a consequence as well as a cause of vowel reduction.

> In his lecture at the workshop, Björn Granström demonstrated

> that this is a distinct possibility: reduction rules applied to synthesized speech resulted in a considerable rate increase even though rate was not directly manipulated. (Engstrand 1992:p340)

Barry (1984) also suggests that the shortening or lengthening of a segment may result from smaller or larger articulatory movements.

Phonology may also take advantage of a phonetic tendency such as increased contextual assimilation as a function of duration. Though phonology may override certain phonetic and physical constraints, it would profit by making full use of natural phonetic tendencies. After all as Lindblom (1983) suggests, economy of speech gestures is an important motivation for the phonology as well as phonetics of languages. In other words, the short duration of schwa may be determined by the interaction of the top-down and bottom-up constraints from the phonology and phonetics of the language.

## 4.4 Conclusion

The degree of variability and the temporal extent of V-to-V effects across a segment for the vowels /ə/ and /æ/ were compared. Schwa was far more variable than the full vowel /æ/. The V-to-V effects were observed right across schwa, while such effects were limited to the transitions or at most the midpoint of the vowel /æ/. The difference in the nature of schwa and the full vowel /æ/ suggests the unique function of schwa in the speech production strategy of so called stress-timed languages. This difference in vowel quality, or more precisely the contrast of targetedness in $F_2$, between full and reduced vowels seems to be an important feature of stress-timing, far more so than has generally been considered in the past. It is also plausible that full vowel in English may be more targeted or hyperarticulated due to this contrast of targetedness compared to vowels in languages where such contrast is not observed.

# Part II

# JAPANESE

# Chapter 5

# The Nature of Japanese Vowels

The purpose of this chapter is to present a brief and up-to-date account of the Japanese vowels. Though some interesting regional differences are mentioned, the description is generally limited to the so-called 'Standard Japanese'.[1] The description of the tone system and accent of Japanese is based on Pierrehumbert & Beckman (1988). It is aimed to clarify the relationship between the intonational structure and the accent of Japanese both of which seem to make an exclusive use of pitch as the material. As the present study focuses on the variation and weakening of vowels, the devoicing process which is an interesting vowel weakening phenomenon in Japanese will be discussed in detail. Finally, some previous studies on vowel variation of Japanese will be reviewed.

## 5.1 Quality

Japanese has a five-vowel system. It consists of five peripheral vowels, /i, e, a, o, u/, each occurring short and long. None are diphthongized. The long vowels can be considered as phonological sequences of two short vowels. Sequences of

[1]This study limits itself to so called 'Hyoojungo' (the 'Standard' Japanese) or 'Kyootsuugo' (the 'Common' Japanese). Kindaichi (1985) prefers the term Kyootsuugo out of sociological considerations, but the two terms refer to the same thing: the language adopted by the Japan Broadcasting Corporation (NHK). It is a near equivalent of refined Tokyo dialect. It is the prestigious common language accepted throughout Japan via media and education. In what follows, unless otherwise stated, when I say 'Japanese,' it refers to Hyoojungo as defined above.

| | | i | e | a | o | u |
|---|---|---|---|---|---|---|
| the present study | $F_1$ | 252 | 418 | 700 | 407 | 307 |
| (5 male speakers) | $F_2$ | 2183 | 1866 | 1178 | 788 | 1173 |
| Imaishi et al (1984) | $F_1$ | 308 | 467 | 775 | 421 | 333 |
| (10 male speakers) | $F_2$ | 2300 | 2039 | 1197 | 708 | 1119 |
| Keating & Huffman | $F_1$ | 359 | 475 | 631 | 481 | 405 |
| (1984) (7 male speakers) | $F_2$ | 1954 | 1720 | 1383 | 1136 | 1419 |

**Table 5.1.** $F_1$ and $F_2$ values of the five Japanese vowels obtained by the present study and two other previous studies.

multiple numbers of vowels are permitted: e.g., *ii aoi e* (a good blue picture), in which case each short vowel is counted as one mora, a rhythmic unit of Japanese to be defined below. In the example above there are six morae in the phrase.

The five Japanese vowels form a skewed pentagon in the $F_1/F_2$ acoustic vowel space due to /u/ which is realized as a fronted unrounded [ɯ][2] as shown in Figure 5.1. The results of the present study are compared with the results obtained by Keating & Huffman (1984). The extent of centralization observed in prose reading in their study is remarkable.[3] Table 5.1 compares the average values of the first and second formants obtained in Experiment 3 of the present study (nonsense words in carrier sentences) with the results obtained by Imaishi *et al.* (1984) (words in isolation) and Keating & Huffman (1984) (prose reading).

Nasalized vowels may also occur in Japanese before the phoneme /N/. /N/ seems to be specified only for the feature of nasality, and its place of articulation is underspecified. It assimilates with the following segment. Thus it appears

[2]Ladefoged & Maddieson (1990) divide one of the major features of vowels, *rounding* into two categories, *compression* and *protruding*. According to Pulleyblank (cited in Ladefoged & Maddieson (1990)), the Japanese vowel /u/ is compressed rather than unrounded. This vowel shows its labiality in alternating with /w/ in verbal inflection; e.g., *iu* (to say) and *iwanai* (the negative of to say). He also notes that an allophone of /h/ which occurs before [ɯ] is bilabial [Φ] with compressed rather than protruded lips.

[3]Though the basic vowel patterns are almost identical, Keating & Huffman's results show a much smaller pentagon. This smaller pentagon is not in the centre of the larger pentagon obtained in the present study, but is situated in the upper part of the larger pentagon. This is because the bilabial context has lowered the $F_2$ values in Experiment 3.

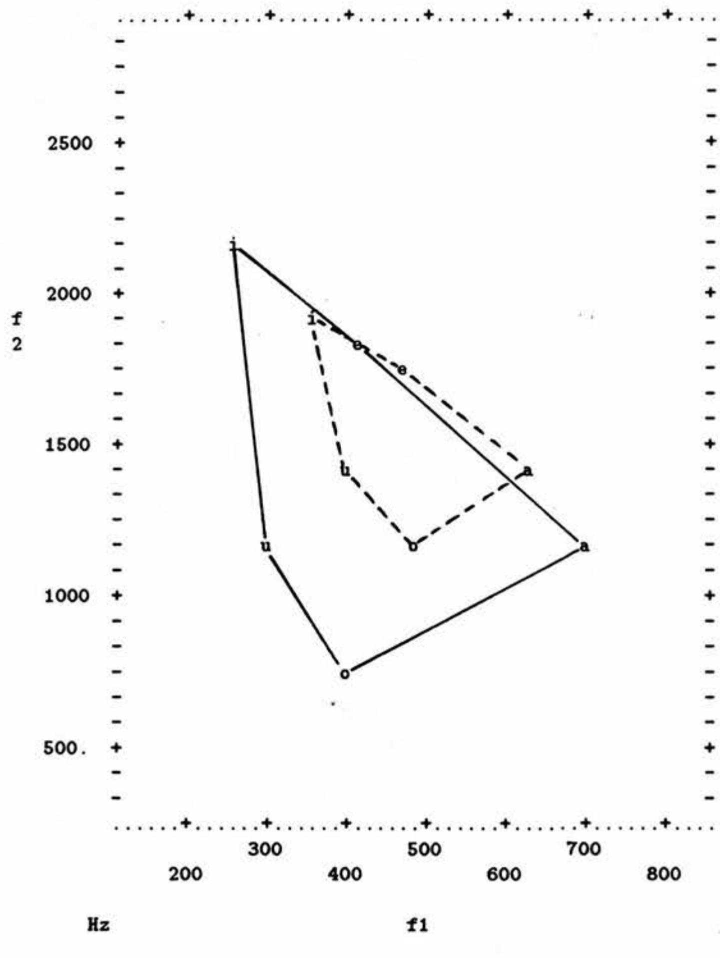



**Figure 5.1.** The mean first and second formant values in frequency Hz for the Japanese vowels /i, e, a, o, u/ obtained by the present study and those obtained by Keating & Huffman (1984). The solid line represents the data from the present study and the broken line represents the data from Keating & Huffman.

as [m] before labial, and as [n] before alveolar, etc.. When /N/ appears word-finally, the nasality seems to appear on the preceding vowel, thus resulting in nasalized vowels. However, nasalized vowels do not have phonological status and are allophonic realizations of the phoneme /N/. This may be a case of gestural overlap between the preceding vocalic segment and the moraic nasal /N/.

Secondary articulation is also observed. There is a phonemic contrast of palatalized vs. non-palatalized consonants before the back vowels /a, o, u/ for the consonants /p, b, t, d, k, g, s, z, m, n, r, h/; e.g., [ka] vs. [kʲa], [ko] vs. [kʲo], and [ku] vs. [kʲu] in words like *kaku* (to miss) vs. *kyaku* (guests).

Devoiced vowels are also observed as free allophonic variants of voiced vowels. Vowel devoicing mainly occurs in high vowels /i/ and /u/, but devoicing has been observed in other vowels as well: e.g., [hḁha] (mother) or [ko̥koro] (heart) (Amanuma *et al.* 1988). Vowel devoicing will be described in detail below.

The distribution of vowel phonemes and their realizations are different from dialect to dialect. In certain areas the high vowels /i/ and /u/ are centralized and realized as [ɨ] and [ʉ]. For example, in the north Oou area, *sisi* (lion), *susu* (cinder) and *susi* (a sort of Japanese cuisine) are all pronounced as [sɨsɨ], and the distinction between the two vowels /i/ and /u/ is lost in certain environments. The appearance of the central vowels seems to be lexically variable. In some dialects, an overlap in /i/ and /e/ is also observed (Imaishi *et al.* 1984). For example, the words *iki* (breath) and *eki* (station) may both be pronounced as [egi] in some dialects.

## 5.2 Quantity

The Japanese vowels have a quantity difference with a distinctive function. For example, for each vowel we may find a minimal or near minimal pair: *i* (stomach), *i'i* (good); *se'ken* (the world), *seeken* (political regime); *obasan* (aunt), *oba'asan* (grandmother); *ko'ko* (individual), *ko'oko* (treasury); *suki* (plow), *su'uki* (strange and mysterious). Some of them are near minimal in the sense that they have different accentual patterns.[4]

[4] The accent is expressed by the diacritic ' after the accented mora.

The short and the long member of each pair are considered to have no qualitative contrast as observed in the English "tense" and "lax" pairs /ʊ/ and /u/ or /ɪ/ and /i/.

According to Crothers (1978) nearly half (45%) of the sample languages in the Stanford Phonology Archiving Project have contrasting long and short vowels. In more than half of the cases (70%) the vowels of the two systems are equal in number and arrangement, either identical in quality or showing only minor differences. Japanese seems to be one of such languages.

According to Lindblom (1963), there seems to be a correlation between vowel duration and the extent of vowel reduction. That is, vowels become more reduced as their length gets shorter. In English, the shorter members of the "tense-lax" pairs are more central in formant values. If vowel reduction occurs as vowels get shorter, there will be some tendency for the short member of the quantitatively contrasting pairs of Japanese to be more reduced than its long counterpart. In some languages this tendency seems to have become enhanced and "phonologized", while in others, the tendency may remain too subtle to be perceptible. The effects of duration on the vowel quality of short and long vowels of Japanese will be discussed in Experiment 3.

## 5.3 Effect of rhythm and melody

In English there is a type of prosodic unit defined by alternations among reduced and unreduced vowels. At a relatively low level in the hierarchy of its rhythmic organization as expressed by a metrical tree, there are stress feet.[5] Contrary to the "rhythm of alternation" as manifested by English, Japanese is said to have a staccato "rhythm of succession". As feet are important measures of English

[5]A stress foot as defined by Selkirk (1980) is a prosodic category which may be a syllable on its own (e.g., the word *gymnast* has two syllables each of which is a stress foot on its own), or it may contain two syllables, strong and weak, as in the word *modest*. A syllable which is a stress foot will never be interpreted as a weak syllable even when the stress foot that dominates it is weak. Thus, being a stress foot always implies some degree of prominence. The term *stress foot* is used by Selkirk to distinguish it from the foot as defined by Abercrombie. A *foot* is defined by Abercrombie (1964) as "the space in time from the incidence of one stress-pulse up to, but not including, the next stress-pulse."

verse, morae are measures of Japanese traditional verses such as "haiku", which consists of 5-7-5 morae and "waka", which consists of 5-7-5-7-7 morae. A mora can be a CV syllable, a single V (a long V is counted as two morae), a "moraic" nasal /N/, or "moraic" first part /Q/ of a long consonant (a long consonant is also counted as belonging to two morae): e.g., *ku.u.ki* (air) = 3 morae; *pa.N* (bread) = 2 morae; *ki.t.te* (stamp) = 3 morae (the first *t* is /Q/).

The role of morae as rhythmic units, however, has been cast in doubt as Beckman (1982) denied the claims made by Han (1962a), Homma (1981) and Port *et al.* (1980) for "moraic" isochrony. Han claimed that when [+high] vowels are "devoiced", voiceless obstruents before deleted vowels are consistently longer than when they are prevocalic, taking up the duration of the omitted vowel to maintain the length of the mora. Han and Homma both cite ratios as large as 3:1 between long and short consonants, which implies that a "moraic first part /Q/" of the long consonant is as long as a CV mora. Temporal compensation has also been cited as evidence for tendency towards "moraic" isochrony (Port *et al.* 1980).[6]

In English, rhythmic units at different levels of the prosodic hierarchy are defined by alternating strong and weak pulses of stress. This may be observed in the alternating **s** and **w** in the different prosodic categories of a metrical tree. Alternations of stressed and unstressed vowels as well as changes in pitch play an important part in organizing speech into units. Such units seem to be also used as perceptual cues. According to Norris & Cutler (1985), the onsets of strong syllables are important cues for speech segmentation in English.

In Japanese, a system defined by accentual phrasing and intonational phrasing, which involves changes in pitch contour, seems to form an organizational unit in speech production. If rhythm may be defined as a recurring pattern, the rhythm of Japanese is marked by its melody, that is, delimitative $F_0$ rise and fall defined by its tone structure. The smallest unit in the intonational system of Japanese is the accentual phrase. The accentual phrase is defined by the phrasal

[6]Fowler (1981) considers that temporal compensation is a correlate of gestural overlap between adjacent segments. Considering vowel devoicing as an overlap in laryngeal (presumably blending) as well as supralaryngeal (hiding) gestures (Jun & Beckman 1993; Browman & Goldstein 1990b), a degree of temporal compensation may be a measure of the extent of gestural overlap observed in vowel devoicing.

H tone and the L% boundary tone which function delimitatively. Following Pierrehumbert & Beckman (1988), accentual phrases are underspecified for tone even at the most surface level. Therefore, there is no one-to-one correspondence between tones and minimal tone bearing units such as syllables or morae as in the traditional account of the Japanese tone structure. The accentual phrase may be either accented or unaccented. The unaccented accentual phrase has a configuration of initial $F_0$ rise to the phrasal H and a slow fall towards the boundary L% tone. When the accentual phrase is accented, the accent HL sequence, which is realized as a sharp fall in $F_0$, is superimposed on the basic delimitative HL% contour. In either case a wave-like rise and fall of $F_0$ characterizes the accentual phrase.

There is a level in phrasing which is intermediate between the accentual phrase and the utterance. This level is referred to as the intermediate phrase (Pierrehumbert & Beckman 1988). The intermediate phrase is the domain of downtrend in $F_0$ known as catathesis. Catathesis is triggered by the accent in the preceding accentual phrase. The effect of catathesis is found to chain. Thus, in a sequence of more than two accented accentual phrases, each subsequent accentual phrase is realized at a lower $F_0$ in a step-like manner until the downtrend is blocked at the reset of tone register at the start of the new intermediate phrase. It should be noted that the same utterance may be realized with different numbers of accentual phrases. That is, for example, the noun phrase *akai se'etaa* (a red sweater) may be realized either as one accentual phrase or two separate accentual phrases.

Even where there is a long series of unaccented phrases, which is possible in Japanese, slightly rising $F_0$ at the onset of each accentual phrase and descending $F_0$ towards the target L% boundary tone gives the language some sort of undulatory rhythm. Furthermore, each subsequent phrase is often realized at a lower $F_0$ though the decline is much less steep than in the cases of accented phrases because of declination which is generally defined as a gradual lowering of the pitch range as a function of time due to phonetic and physiological constraints. Thus, Japanese is characteristically defined by the rise and fall of $F_0$, induced by phrasing and accents, creating a sort of wave-like melody. Large waves of intermediate phrases contain smaller accentual phrasal waves.

## 5.4 Accent

In defining accent, I follow Beckman (1986). The accent as defined below may be divided into two types of accent: stress-accent such as represented by English and non-stress accent which may be represented by Japanese. Phonetically stress-accent uses material other than pitch to a greater extent than non-stress accent does. Accent is,

> a system of syntagmatic contrasts used to construct prosodic patterns which divide an utterance into a succession of shorter phrases and specify relationships among these patterns which organize them into larger phrasal groupings.
>
> it (accent) makes one part of a word or a phrase culminatively more or less prominent than the rest, and in so doing marks out the organizational unit in an utterance. (Beckman 1986:p1)

Accent in Japanese seems to play a less important role in its culminative function compared to, for example, English stress-accent, in that long sequences of unaccented phrases are permissible. For example, it is possible to have the whole sentence without any word accent in Japanese: e.g., "Kimi-no tomodachi-no Tanaka-san-no shigotoba-wa koko-da" (Your friend Mr. Tanaka's office is here). Furthermore, it has been noted that 28,031 words (51.2%) out of 54,187 words listed in the Sanseido Dictionary have turned out to be unaccented words (Kubozono 1987). In Japanese, it seems that the undulatory rise and fall of $F_0$ defines the prosody of the language. Accent seems to have only a minor role in giving prominence to certain parts of utterances by creating a larger pitch wave contour.

The following description of the Japanese pitch accent is based on Pierrehumbert & Beckman (1988). The primary phonetic feature of accent in Japanese is a sudden drop in $F_0$. There is at most one accent per word or phrase. The accentual HL fall is much steeper than the delimitative HL% contour in an unaccented accentual phrase. When an accentual phrase is unaccented, there is a decline in $F_0$ from the phrasal H to the boundary L%. This decline is realized almost as a straight line interpolated through from the phrasal H to the boundary L%.

This slope becomes less and less steep as the number of morae between the two target tones increases. There seems to be a certain target value for the L% tone to which the $F_0$ decline aims. There is also strong and weak allophony of the L% tone. That is, when the first syllable of a phrase is accented or long, the L% tone of the preceding phrase is weak in a sense that it does not receive the full duration and low $F_0$ value that characterize the L% tone before an unaccented short syllable.

When the accent HL sequence is superimposed on this basic delimitative contour, two lines intersecting at the accent L tone may be observed on the pitch configuration. The first line is a steep line interpolating through the accent H to the accent L, and the second line is a shallow slope interpolating through the accent L to the boundary L%.

The unit to which accent is assigned seems to be a syllable. The accent H seems to align with the first mora of the accented syllable. The location of the accent is specified in the lexicon and not predictable, but given the location of the accent, other pitch features of words and phrases are predictable.

Accented words and phrases invariably have a higher $F_0$ peak than unaccented phrases. Not only the peak but the trough between the two phrases is higher before an accented phrase (weak allophony). Thus, the effect of accent seems to spread leftward (Kubozono 1987). Accented phrases tend to show a greater degree of initial rise irrespective of the location of the accent within the phrase. When the accent H occurs relatively late in a phrase, we may see two different pitch heights of the phrasal H and the accent H realized in the pitch configuration of one accentual phrase (Pierrehumbert & Beckman 1988). The overall higher value of $F_0$ in accented accentual phrases is called 'accentual boost' by Kubozono (1987).

Accent triggers catathesis, lowering everything to the right of the accented accentual phrase. Accent H's, phrasal H's, terminal boundary H%'s in questions as well as the L% boundary tone and the L within the triggering accent are all affected. Other things being equal, they are lower in $F_0$ when they follow an accented phrase than when they follow an unaccented one.

While stress accent languages use material other than pitch, such as duration, amplitude and vowel quality as phonetic cues of accent, accent in Japanese seems to use pitch as its only main phonetic correlate (Beckman 1986). Though pitch seems to be the strongest cue in stress accent languages as well (Fry 1958), Nakatani & Aston (1978, cited in Beckman 1986) showed that the relative effectiveness of the $F_0$, duration and vowel quality varied depending on the position of the word in the sentence. When the word had nuclear stress, the $F_0$ pattern outweighed the other parameters. In prenuclear positions, duration and vowel quality vied with $F_0$, and in postnuclear position, duration outranked $F_0$. On the other hand, in an experiment to test the perceptual cues to accent in English and Japanese, Beckman (1986) showed how extensively and exclusively the fundamental frequency was used to interpret the Japanese stimuli. She used three groups of subjects: Japanese, American bilinguals of Japanese and American monolinguals. In interpreting the Japanese stimuli, the mean effect-on-accent scores for the fundamental frequency was higher for the Japanese subjects than for the American bilingual subjects, who in turn had higher scores for the $F_0$ than did the American monolinguals. The scores for the duration, amplitude and spectral pattern did not exceed the chance level to any significant degree. (See page 225 for more discussion.)

Beckman (1986) also showed that fundamental frequency is the only major acoustic correlate of accent in Japanese. She compared the patterns of the three variables, fundamental frequency, amplitude and duration in disyllabic minimal pairs contrasting in accent, such as *pérmit* (noun) vs. *permít* (verb) and *i'ken* (opinion) vs. *iken* (different view). She compared the values for these variables between the first and the second syllable nucleus within each word of each language. She observed remarkably similar patterns for fundamental frequency between English and Japanese. On the other hand, the patterns for duration and amplitude showed marked differences between the two languages. The results as a whole suggested that while amplitude and duration were strongly correlated with accent in English, there is little correlation between accent and these variables for Japanese. Thus, fundamental frequency seems to be the major perceptual cue and acoustic correlate of accent in Japanese.

## 5.5 Vowel devoicing

In Japanese, vowels, especially high vowels /i/ and /u/, tend to shorten, devoice or disappear between voiceless consonants [p, t, k, Φ, s, ʃ, ç, h, ts, tʃ] or in an unaccented word final syllable such as [ki, kɯ, ʃi, sɯ, tʃi, tsɯ, çi, Φɯ] (Hirayama 1985).

This phenomenon is considered to be free allophonic variation and the extent of vowel devoicing is considered to vary from dialect to dialect. In the Southern RyuuKyuu dialect, the low vowel /a/ as well as high vowels devoices: e.g., /pḁnna/, (flower). On the other hand, in Kyoto, Shikoku, Chuugoku, and Chuubu areas, where vowels are articulated more carefully, vowel devoicing is not observed so commonly (Hirayama 1985).

The sociolinguistic conditions for the variation are not well understood, but a moderate amount of devoicing seems to be a mark of the prestigious Standard Japanese. Hirayama (1985:p44) advises in Nihongo Hatuon Akusento Ziten published by the Japan Broadcasting Corporation that "though overmuch devoicing is not preferable, appropriate devoicing improves the coherence of words and phrases, giving a feeling of articulate crispness to one's speech. Thus, the extent to which devoicing is observed in the contemporary standard language is rather appropriate and recommendable."

The traditional phonological description of devoiced vowels is that they are voiceless variants of /i/ and /u/ ([i̥] and [ɯ̥]) accompanying no laryngeal adduction or voice, but with the lingual configurations for the vowels /i/ and /u/ (Kawakami 1977). Beckman & Shoji (1984) also describe the phenomenon in the following manner,

> the vowel as a laryngeal gesture towards the goal of adducted vibrating vocal folds begins later than the vowel as a lingual gesture. When the vowel is deleted for a devoiced syllable, then, it may be only the time portion associated with the vowel as a laryngeal gesture that is removed. A portion of the lingual gesture, if it is compatible with the gesture towards the following consonant, may be maintained.

Acoustically, however, there is often no sign of a devoiced vowel. The waveforms or spectrograms typically show the frication noise associated with the preceding consonant. Formant-like bands indicating the presence of voiceless vowels are often lacking. Furthermore, duration measurements show little difference between the length of a devoiced syllable and that of the consonant in a syllable where devoicing has not taken place, even between different tokens of the same word produced by the same speaker as free variation (Beckman 1982).

Yet the contrast between the members of a minimal pair such as [ʃi̥kan] and [ʃɯ̥kan] is maintained. Beckman & Shoji (1984) recognize the significantly lower mean frequency of the fricative's noise band in the tokens where /ʃu/ is intended. The consonant preceding the devoiced vowel seems to coarticulate with the underlying deleted vowel. The results of the perception test conducted by Beckman & Shoji showed that the identification rate of the words in minimal pairs containing devoiced /ʃi/ or /ʃu/ was 73%, which is well above the chance level. This result was not much different from the mean 65% correct-response rate that Yeni-Komshian & Soli (1981) obtained for vowel identification in [ʃ]'s excised from /ʃi/ and /ʃu/ syllables in English. Moreover, the colouring observed in the Japanese tokens is similar in magnitude to the 200Hz difference that Soli (1981) found for the lowest spectral peaks in the last 60 ms of the fricative in /ʃi/ versus /ʃu/ in English. The anticipatory coarticulation of /ʃ/ with the vowels /i/ and /u/ is comparable between English and Japanese. (See also 8.2.4: p 205.)

Keating & Huffman (1984:p196) report a number of cases where vowels were shortened in the environment where devoicing is expected. They claim that

> deletion and shortening appear to be two degrees of the same phenomenon, so that overall, no vowels were uniformly absent from particular contexts. Rather, in many repetitions of any given word, a vowel will sometimes be deleted, sometimes be extremely short, and sometimes be neither.

As a whole vowel devoicing seems to be a result of gestural overlap (Jun & Beckman 1993). A devoiced vowel seems to be subsumed under the strong frication of the preceding obstruents. Depending on the extent of overlap, it is sometimes shortened and sometimes deleted.

Devoicing may occur on accented syllables as well. This is an interesting case in the phonology because pitch accent cannot be realized when the accent carrying syllable is devoiced, for pitch is absent on that syllable. One strategy for recovering the pitch accent on a devoiced vowel is to shift the pitch drop to the right to fall on a voiced segment (Lovin 1976). The shift seems to have attained the status of phonological rule at least in some cases (McCawley 1968; Akinaga 1985). For example, the adverbial form of the accented three-mora adjective has an initial accent as a rule as in *a'oku*, *shi'roku* and *ta'kaku*. However, when the first mora contains a devoiced vowel, the accent is shifted to the next mora to the right as in [tʃi̥ka'kɯ], [hɯ̥ka'ku] (Akinaga 1985). Akinaga, however, states that there is a trend among the younger people to pronounce the accented devoiced syllable without shifting the accent as in [hi̥'sako] (a girl's name).

A number of experiments on the state of larynx during the devoiced syllable have been conducted at the Research Institute of Logopedics and Phoniatrics at the University of Tokyo, using electromyographic (EMG) and fiberoptic techniques. In the case of CV̥C sequences where the C's are voiceless, a single glottal opening with a long duration and a large maximum glottal width was observed. Different voiceless consonants have different maximum glottal widths. The opening of the glottis for the CVC sequence containing a devoiced vowel was generally larger than or comparable to that for voiceless consonants including geminate consonants (Sawashima 1971).

Thus, the opening of glottis during a devoiced vowel in Japanese seems to extend to its full length. Devoicing is presumably a sort of laryngeal coarticulation. However, it may not be merely a passive act, as reported by Sawashima (1971:p13),

> the glottal adjustments for devoicing of the vowel are not a mere skipping of the phonatory adjustments for the vowel but a positive effort of widening of the glottis for the devoiced vowel segment, even though there is no phonemic distinction between the voiced and devoiced vowels.

The results of the EMG studies show difference in the activities of the abductor (Posterior Cricoarytenoid=PCA) and adductor (Interarytenoid=INT) muscles between the voiced and devoiced vowel (Sawashima *et al.* 1978). The activities of Lateral Cricoarytenoid and Vocalis muscles, which perform the finer adjustments in glottal adduction, have also shown differences between the voiced and voiceless free variants of a vowel (Hirose *et al.* 1970). Sawashima *et al.* (1978) observed two peaks in the activity of the PCA muscles in some of the words produced with a devoiced vowel: the words /si̥see/ and /si̥ssee/ produced by the speaker HH and the word /si̥ppee/ produced by the speaker MS. It is worth noting that though the glottal opening was realized as one continuous gesture, there were two peaks in the muscle activity. (See also Yoshioka (1981).) The double peak activity of the PCA muscles may be interpreted as a result of the speaker's effort to produce a wider glottal opening. It may also be interpreted as a result of the overlap of two separate glottal opening gestures. (See below.)

There are cases when contextually induced vowel devoicing occurs in whispered speech. A clear abducting gesture in which the glottal width had increased from a relatively narrow constriction of whisper was observed corresponding to the time of devoiced syllables in whispered speech (Weitzman *et al.* 1976). Whispered vowels are classically defined as those produced with a narrowing (or even closing) of the membranous glottis while the cartilaginous glottis is open. A triangular opening of the cartilaginous glottis is a physiological feature of whisper. Air flowing through the narrow constriction produces the turbulent hissing sound characteristic of whisper. Catford (1964:p31) describes the aerodynamic and acoustic aspects of whisper as 'turbulent flow, with production of high-velocity jet into pharynx: Acoustic spectrum similar to breath but with considerably more concentration of acoustic energy into formant-like bands'. Whisper characterizes the utterance-final devoicing process observed in many languages, including English and French (Laver 1980). Keating & Huffman (1984) report that true voiceless vowels in Japanese occurred only at the ends of utterances, when one or more entire syllables were devoiced, and this utterance-final devoicing affected all vowels. It is plausible that devoicing observed at the ends of utterances is essentially different in nature from the vowel devoicing process observed at the other positions of the utterance.

In summary, a so-called vowel devoicing process in Japanese may involve two different phonetic phenomena, utterance final vowel devoicing (whisper) and the gestural overlap of a vowel with the preceding voiceless obstruent.

Browman & Goldstein (1990b) discuss two types of gestural overlap, those within and across tiers. A gestural overlap across tiers occurs when the articulators involved are different. For example, in connected speech, the sequence *must be* is often realized as [mʌsbi]. The tongue gesture for [t] is acoustically *hidden* under the labial gesture of [b]. That is, the two gestures completely overlap with one another, giving an impression to the listeners that [t] has been deleted. A gestural overlap within a tier happens when the two gestures share the same articulator. For example, in the sequence *ten things*, the /n/ is realized as [n̪] as a result of the *blending* of the gestures of [n] and [θ]. In vowel devoicing both types of gestural overlap are involved, one at the supralaryngeal and the other at the laryngeal level.

At the supralaryngeal level, the tongue body is relatively free during the articulation of the preceding obstruent (Keating 1988; Carney & Moll 1971). Thus, the vowel gesture may be acoustically hidden within the preceding segment. On the other hand, at the laryngeal level, the gesture for the preceding voiceless obstruent [+ open glottis] and the gesture for the vowel [− open glottis] seem to be incompatible. What seems to be happening is not an intermediate degree of glottal opening, but a clear cut [+ open glottis].

Vowel devoicing seems to be favoured when the preceding segment has a strong frication period which requires a wide opening of the glottis (Kagaya 1974). Fricatives and affricates are reported to induce a higher rate of vowel devoicing than plosives (Han 1962a; Takeda & Kuwabara 1987; Yoshida & Sagisaka 1990; Kondo 1993; Jun & Beckman 1994).

On the other hand, vowels are more likely to be devoiced before stops than before fricatives (Takeda & Kuwabara 1987; Kondo 1993; Jun & Beckman 1994). Jun & Beckman (1994) suggest that glottal vibration may cease sooner in a vowel-stop sequence than a vowel-fricative sequence as the quick buildup of oral air pressure needed for the production of a stop seems to be associated with faster and more ballistic articulatory movements into stops than into sibilants. Thus, the occurrence of vowel devoicing seems to be motivated by a number of physical constraints.

Jun & Beckman (1993) argue that the same-tier overlap is possible. The glottal width data from Munhall & Löfqvist (1992) show that in tokens of the phrase *Kiss Ted* spoken at different tempi, two separate opening and closing movements for the consonants /s/ and /t/ were observed at slow tempi while at faster tempi, only a single glottal gesture was observed for the /s/-/t/ sequence (Figure 5.2). The EMG data from Yoshioka (1981) also supports the same-tier overlap of the glottal gesture. In different tokens of the word /hisee/, two peaks for the activity of the abductor muscles (PCA) were observed when the vowel /i/ was voiced. On the other hand, when the vowel /i/ was devoiced, there were cases where there was a single peak of the PCA activity as well as cases where two peaks seemed to have come so close together that they constituted a single gesture (see Figure 5.3).

To conclude, the gestural overlap account of vowel devoicing seems to be a simple and clear description of the observed phenomena. Further, Jun and Beckman argue that vowel devoicing may be a universal vowel weakening phenomenon that may exist in most languages of the world (see page 269). In English schwa is the target of this process (see page 66). Vowel devoicing and vowel reduction seem to be two different processes defined at different levels of the linguistic representation. In my view, while vowel reduction is determined by the phonology of the language, vowel devoicing seems to be defined by sociolinguistic and pragmatic as well as phonetic motivations. See page 269 for more discussion.

## 5.6 Vowel variation

Keating & Huffman (1984) have observed that Japanese vowels vary to a much greater extent in prose reading than in word list reading. The vowel pattern of Japanese is such that there is an empty region in the high, back area of the available vowel space. This and other peripheral regions remain clearly empty and unfilled in prose reading. Vowel variation does not seem to explore these peripheral empty regions. It occurs mainly in one direction. That is, vowel variation spreads inward, filling the centre, sometimes causing overlap and perceptual confusion between different vowels. The basic vowel pattern remains the

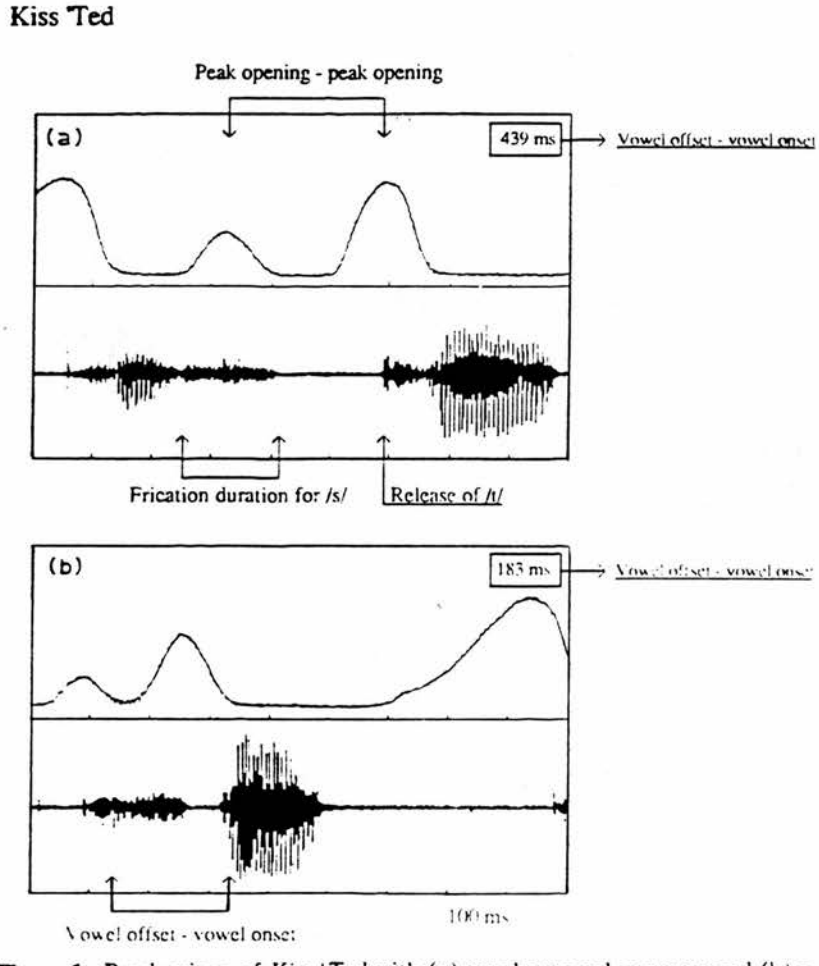



**Figure 1.** Productions of *Kiss 'Ted* with (a) two laryngeal gestures and (b) a single laryngeal gesture for /s/ and /t/. Selected articulatory events and intervals are indicated.

**Figure 5.2.** Figures taken from Munhall & Löfqvist (1992).

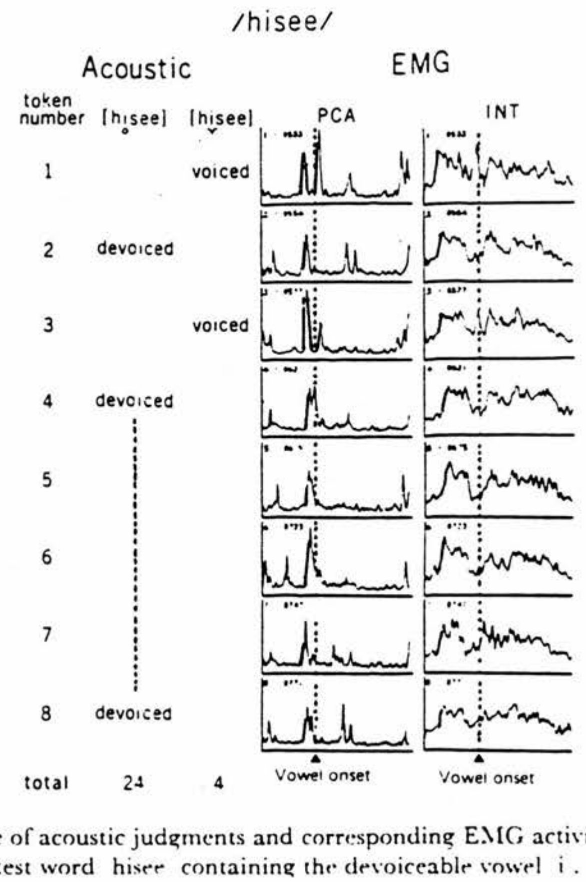



*Fig. 3.* Sample of acoustic judgments and corresponding EMG activity patterns of PCA and INT for the test word hisee containing the devoiceable vowel i .

**Figure 5.3.** Figures taken from Yoshioka (1981).

same skewed pentagon and allophonic variation seems to be confined within the basic phonemic inventory. This is presumably a result of the greater extent of coarticulation in prose reading.

Kuwabara (1972) has observed a large magnitude of vowel variation in the middle vowel placed between two adjacent vowels in three-vowel sequences such as *tasuk***eae***ba* (if (we) help one another). There were 20 symmetrical sequences such as [iai] and 20 asymmetrical sequences such as [iae] placed in short sentences. Figure 5.4 shows a plot of Kuwabara's data. It is clear from the figure how each vowel is pulled towards the value of its adjacent vowels in the symmetrical sequences. The four realizations of the vowel /e/ between /i/'s, /a/'s, /o/'s and /u/'s are the clearest example. The pattern obtained by linking these points is almost the same as the basic vowel pattern of the phonemic inventory. The difference between the values of /e/ between the front high vowel /i/'s and the low vowel /a/'s is about 175Hz in $F_1$. The difference between the values of /e/ between /i/'s and the back vowel /o/'s is about 200Hz in $F_2$. Though the effect of surrounding vowels is expected to be less when they are non-adjacent, variability of vowels due to contextual effect is striking, particularly in $F_1$. It is also interesting to note that the distribution of the vowels in the language seems to affect the pattern of vowel variation (Manuel & Krakow 1984; Manuel 1990). The front vowels /i/ and /e/ are clearly separated from the back vowels /a, o, u/ in the figure. The front vowels show a spread in $F_1$ while the vowel /a/ mainly extends in $F_2$. The back vowels /o/ and /u/ spread both in $F_1$ and $F_2$. These vowels seem to spread in a way that avoids extensive overlap with adjacent vowels.

Magen (1984) reports that in a pilot study on Japanese VCV utterances, the magnitude of coarticulation was observed to be greater in Japanese than in English. She compared the V-to-V coarticulation in English and Japanese using VbV sequences with the vowels /i/ and /a/. The data included two possible stress patterns in English (V́bV and VbV́) and two pitch patterns in Japanese (HL and LH). When the second formant trajectories across the vowel as a function of the preceding (or following) vowels, /i/ or /a/, were compared, the V-to-V coarticulatory effects were stronger on unstressed vowels than on stressed vowels in English. When the $F_2$ trajectories were compared between the English

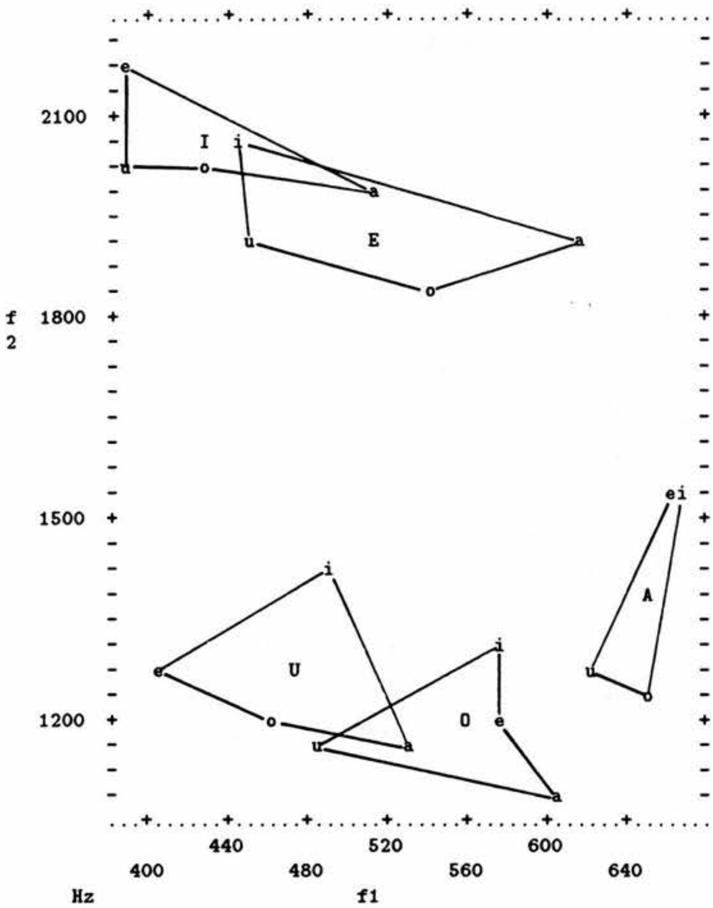



**Figure 5.4.** The V-to-V coarticulation data from Kuwabara (1985). The symbols 'I, E, A, O, U' represent the mean $F_1$ and $F_2$ values across all contexts. The symbols 'i, e, a, o, u' around the grand means show the mean formant values in the symmetric contexts of /i/, /e/, /a/, /o/, /u/.

unstressed vowels and the Japanese vowels, the effects of the transconsonantal vowels were stronger on Japanese vowels. The V-to-V effects were observed right through the duration of the vowel in Japanese, while in English, the effects tended to diminish towards the end of the segment for carryover effects, and the effects were not clear at the beginning of the segment for anticipatory effects. In particular, anticipatory effects were observed to be very strong in Japanese. From these it seems that Japanese has rather extensive vowel variation due to contextual effects. Possible sources of vowel variation in Japanese including contextual effects will be explored in the experiments to follow.

# Chapter 6

# Vowel Variation in Japanese

## 6.1 Introduction

A series of experiments was conducted to explore possible sources of vowel variation in Japanese. These experiments are presented in Chapters 6 through 8. There are two main objectives in carrying out these experiments.

First, a preliminary survey of the nature of vowel variation in Japanese is intended in these studies. To my surprise, there have been very few systematic studies on variation and the coarticulatory pattern of the Japanese vowels. The few studies I have come across are Kiritani *et al.* (1977), Keating & Huffman (1984), Magen (1984), Kuwabara (1972) and Beckman (1986). The purpose of the present study is to supplement these, and to provide more data on the effects of variables that may affect the vowel quality of Japanese. In this sense, the present study is descriptive in nature.

Secondly, the coarticulatory pattern of Japanese, a non-stress accent language, will be studied in comparison to the coarticulatory pattern of English, a stress accent language. In Experiment 1, the coarticulatory pattern of English schwa was observed, and it was concluded that schwa may be targetless in $F_2$. The results suggested that vowel reduction may be contextual assimilation rather than centralization. In Experiment 2, the variability and transparency of schwa were compared with those of the full vowel /æ/. Schwa was observed to be far more variable than the vowel /æ/. Schwa was also more transparent than the vowel /æ/ in that V-to-V effects were observed right through the schwa (often from the onset to the offset, and vice versa), while such effects were stopped at the midpoint or did not go beyond the transition for the full vowel /æ/. It was suggested that the contrast of 'targeted' and 'targetless' vowels and the transparency of schwa may be important characteristics of stress-timing.

On the other hand, Japanese is a non-stress-timed language. Its coarticulatory pattern must be essentially different from that of English. Though the exact nature of its coarticulatory pattern may not be predicted, certain hypotheses may be presented to speculate its nature.

Firstly, there have been no reports of difference in the extent of accent related vowel variability among the five vowels of Japanese apart from the devoicing of two high vowels /i/ and /u/. In this sense, the two [+high] vowels may contrast with the [-high] vowels in Japanese. However, they are often deleted before they could show the effects of the surrounding segments on them. At this stage, I would like to treat this process of vowel devoicing separately from vowel variation. For further discussion on the theoretical issues involving the treatment of vowel reduction, devoicing and weakening in general, see page 269 in the general discussion.

Apart from high vowel devoicing, there is no report that Japanese vowels have a contrast of targeted and targetless vowels such as the one observed between the full and reduced vowels of English. On the other hand, greater extent of variability was observed (both in magnitude and temporal extension) for Japanese vowels than for English full (stressed and unstressed) vowels (Magen 1984). In order to enhance the contrast in the extent of context dependent vowel variability, English full vowels in general may be more hyperarticulated than vowels in languages where no such contrast is observed. From these observations, the following hypothesis is proposed.

- Having no contrast of targeted and targetless vowels, the degree of vowel variability in Japanese may be intermediate in degree between that observed for the English schwa and the English full vowels.

In English, stress accent is an important factor in introducing the contrast of targeted and targetless vowels. If vowel reduction is largely a result of contextual assimilation as suggested in Chapter 3, English stress would play an important role in determining the extent of vowel variation as a function of context. On the other hand, very little correlation was observed between accent and vowel quality for Japanese (Beckman 1986). Accent seems to be minimally related to vowel variability in Japanese. From these observations. the following hypothesis is proposed.

- Accent does not affect the extent of context dependent vowel variability in Japanese.

In the present study, the above two hypotheses will be tested in order to illustrate the features of the coarticulatory pattern of Japanese that contrast with that of English.

The main variables considered in the present study include accent, pitch, vowel duration (phonemic contrast), syllable position (initial and final in disyllabic words) and contexts (both consonantal and vocalic). Some of the variables treated here are generally considered to have little effect on vowel variability of Japanese. However, in order to present a more comprehensive description of the nature of vowel variation in Japanese, even minor effects (the effects that may not be linguistically significant at the present stage but may have potential to be significant) were considered. Some factors that are not covered here are speech rate, style, type of task (e.g., word list reading, prose reading, etc.).

The above variables were studied in a number of different experiments using different materials. The effects of accent, pitch, vowel duration and syllable position were observed on all the five vowels /i, e, a, o, u/ of Japanese. The effects of consonantal and vocalic contexts were observed only for the vowels /e/ and /a/. The variability of the vowels /i, o, u/ as a function of contexts were not studied in the present study. There may be inherent variability for each individual vowel. For example, the high vowels /i/ and /u/ that are inherently shorter (Lehiste 1970) and smaller in amplitude (Lehiste & Peterson 1959) may be more susceptible to variability. In other words there may be difference in variability as a function of vocalic features, for example, front vs. back, high vs. low and rounded vs. unrounded (Stevens & House 1963). This issue was not addressed in the present study, but obviously it is an interesting issue to be addressed in future.

The contextual segments considered in the present study are also limited. All the five vowels /i, e, a, o, u/ were considered as contextual vowels. The consonants /p, t, k/ and the /s_k/ context were used as consonantal contexts. In order to observe the effects of secondary articulation on V-to-V coarticulation, the /b, $b^j$/ contexts were also studied.

In summary, the present study does not cover all the aspects of vowel variation in Japanese. Such comprehensive coverage of the topic is beyond the scope of the present dissertation. Some of the data presented here are also limited in the number of tokens and subjects, and may only serve as a preliminary survey for future studies. However, they are suggestive of important features of the coarticulatory pattern of Japanese and therefore are presented in this dissertation.

## 6.2 Experiment 3: A preliminary survey

An experiment was conducted to study vowel formant variation as a dependent variable of accent, pitch height, vowel length, and syllable position (initial and final in disyllabic bVbV and bVVbVV nonsense words). The variables studied here are generally considered to have small effects on the vowel quality of Japanese. However, as stated earlier, the nature of the present experiment is primarily descriptive. It is intended to provide a detailed description of the nature of vowel variation in Japanese.

### 6.2.1 Methods

#### Speakers

Seven male subjects were recruited for the experiment. They are SE, KF, JK, TK, YK, KO and KM. All of them have lived most of their life in or near Tokyo and consider themselves to be 'Standard' Japanese speakers. They have also had some experience of living abroad, from a few months to about three years. At the time of the recording they were either living in or visiting Edinburgh, U.K.. KM, who had lived in Kyoto, a town in the Western part of Japan, till the age of ten, had an obvious Kyoto accent in his recording, though not so noticeable in his ordinary speech that his recording had to be discarded. JK and TK were born and have grown up in Tokyo. SE is from Chiba, a prefecture to the East of Tokyo. JK and SE had a problem in producing nonsense word tokens. SE's recording had to be discarded. JK made a second recording. KF is from Kanagawa, a prefecture to the West of Tokyo. YK is from Saitama, a prefecture to the North of Tokyo. KO has lived in Tokyo since the age of five. In the end, it was possible to use the recordings of KF, JK, TK, YK and KO for analysis.

### Materials

The test material consists of disyllabic nonsense words in bVbV and bVVbVV sequences (VV = a long vowel). These words may be divided into four types according to their tonal and accentual patterns. Originally they were intended to be (1) CV'CV with an initial accent[1], (2) CVCV' with a final accent, (3) CV'VCVV with an initial accent, and lastly (4) CVVCVV without accent, but as will be described in detail below, the type (2) words were realized as unaccented in most cases. All the five vowels /i, e, a, o, u/ were used. These test words were presented to subjects in Japanese orthography *katakana* (straight syllabary) with accentual diacritics used by Nihongo Hatuon Akusento Ziten (NHK ed., 1985). The accent is marked by a corner on the accented mora and the unaccentedness is indicated by a straight line over the transcription starting from the phrasal H (the second sonorant mora) to the end of the word as shown below: However, since most subjects are not familiar with such diacritics, the accented or High toned syllables were emphasized in bold type. All the test words were placed in the frame sentence: *Kore wa ___ desu* (This is ___).

| | | |
|---|---|---|
| kuzu | rubbish | accented on the first mora |
| kudamono | fruit | accented on the second mora |
| kusuribako | medicine box | accented on the third mora |
| kudarizaka | downhill slope | unaccented |

[1]The diacritic ' indicates that the accent falls on the syllable preceding it.

The type (2) words are problematic as mentioned above because finally accented words and unaccented words of the same mora length are neutralized in isolation. For example, the words *hana'* (flower) (finally accented) and *hana* (nose) (unaccented) do not contrast in isolation. It is only when they are followed by *desu* (copula) with which they constitute a larger accentual phrase, for example, in the sentence *Kore wa hana' desu* (This is a flower) or *Kore wa hana desu* (This is a nose) that the two words are distinguished by the accentual fall that marks the accent in *hana'*. Thus, the type (2) nonsense words have potential to be interpreted as either finally accented or unaccented, and despite the accent diacritic placed at the end of the word, in most cases they were realized as unaccented. Two subjects KF and JK read the type (2) words as accented. The other subjects read the words as unaccented. Figure 6.1 shows $F_0$ contours of the sentence *Kore wa baba desu* produced with and without accent by the subjects KF and KO respectively. As KF laryngealizes towards the end of the utterance, $F_0$ contour is not observed for the portion of the copula *desu*. Yet the hint of accentual drop may be observed on the second vowel of the word as the $F_0$ contour starts to fall in a rather steep slope. For the words of the CVVCVV type with two long vowels, there are only two possible accentual representations, i.e, either initially accented or unaccented. The following is a list of nonsense words used in this study. They were repeated five times in a random order.

| (1) | (2) | (3) | (4) |
|---|---|---|---|
| ba' ba | ba ba | ba'a ba a | ba a ba a |
| bi' bi | bi bi | bi'i bi i | bi i bi i |
| bu' bu | bu bu | bu'u bu u | bu u bu u |
| be' be | be be | be'e be e | be e be e |
| bo' bo | bo bo | bo'o bo o | bo o bo o |

$F_0$ contours of the four types of nonsense words produced by TK are given in Figure 6.2. Higher $F_0$ peak and a sharp accentual drop are observed for both type (1) and (3) words. For the type (2) and (4) words that are unaccented, $F_0$ stays rather level after the initial rise and declines very slowly. The peak $F_0$ is lower than that of the accented word.

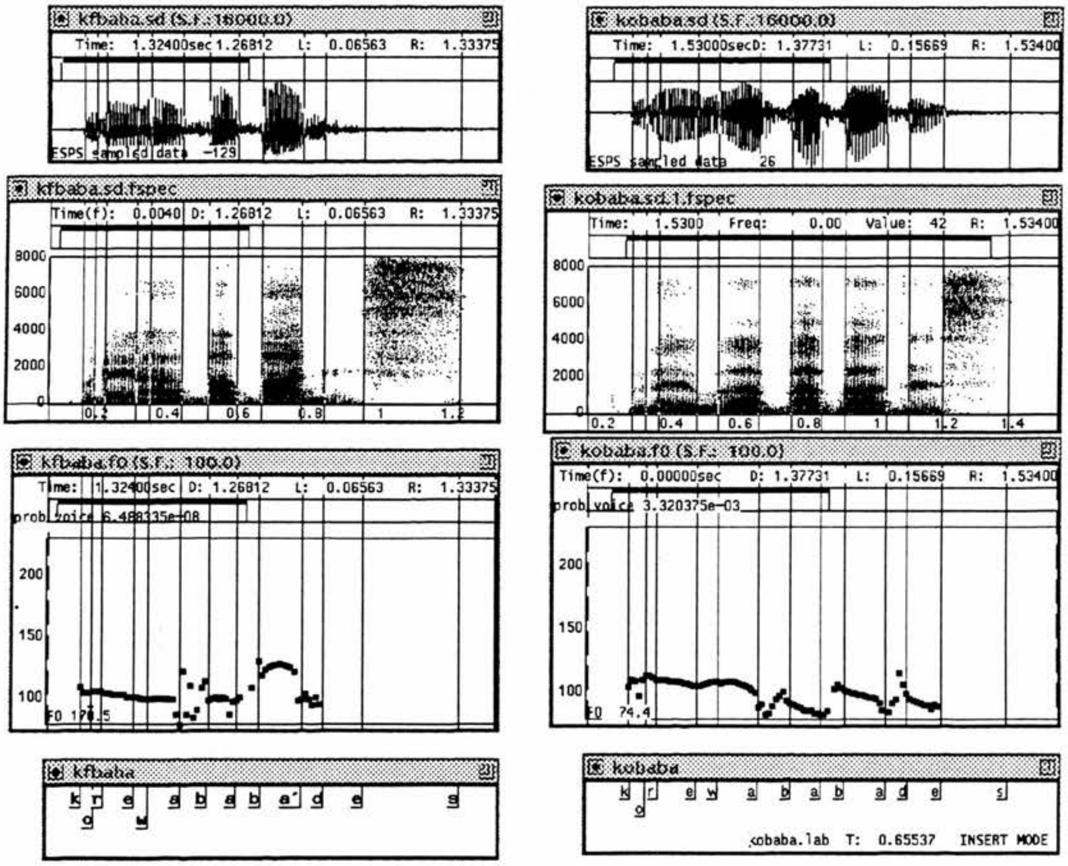



**Figure 6.1.** The $F_0$ contours of the utterances 'Kore wa baba desu' with and without accent on the final syllable of the nonsense word 'baba' produced by KF (on the left) and KO (on the right).

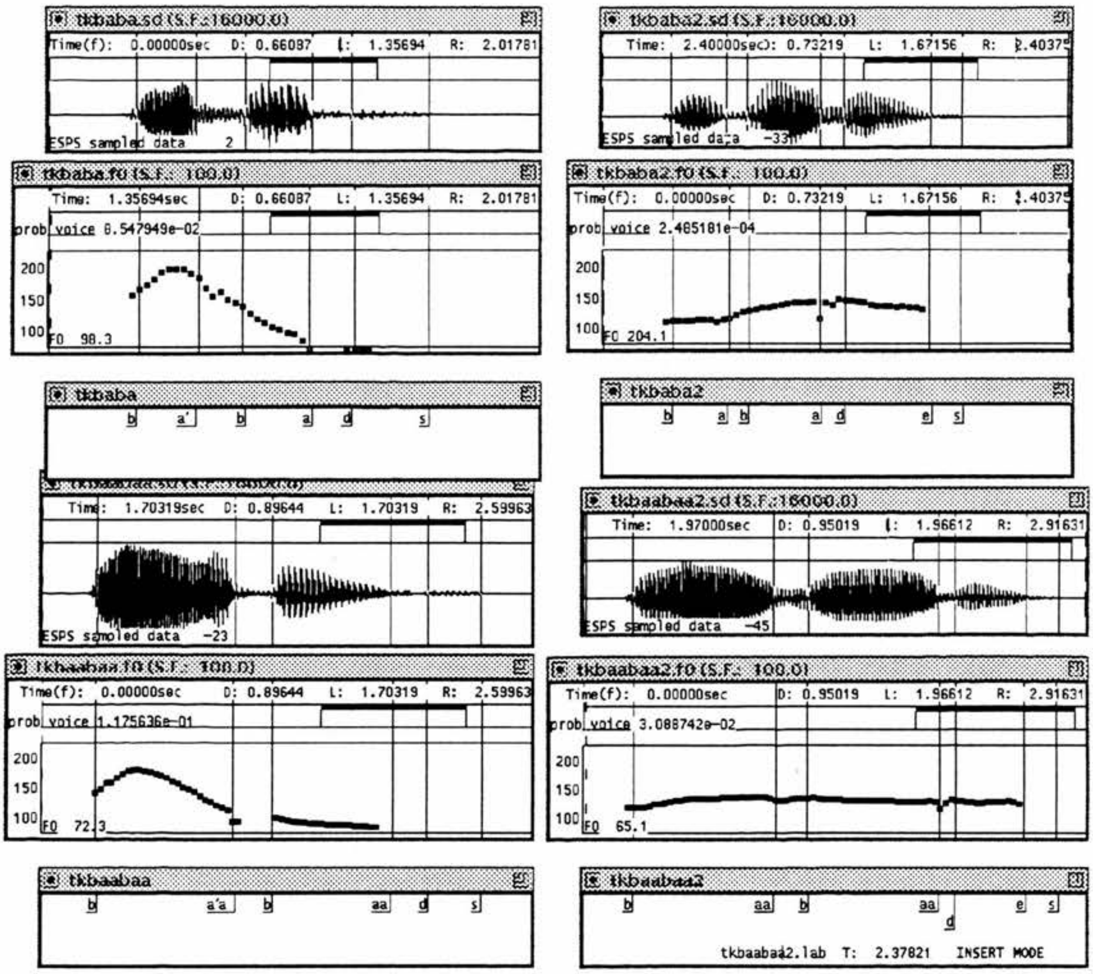


**Figure 6.2.** The $F_0$ contours of the four test words ba'ba, baba, ba'abaa and baabaa produced by TK.

The total of 100 tokens (20 test words $\times$ 5 repetitions) were produced by each subject. For the analysis, only the middle three repetitions were used unless there was some problem. In such cases tokens from either the first or the last repetition were used instead. The total of 60 nonsense words were analyzed for each subject. These tokens include 120 vowels for analysis for each subject.

### Recordings

Recording was done in a sound treated recording studio equipped with a Sennheiser MKH815T microphone with RF condenser connected to Soundcraft 200B 8-4-2 audio mixing console through which the sound was passed into a Sony PCM 701 ES digital audio processor and a Sony SLF 25 Betamax videorecorder and to a Ferrograph Logic 7 stereo tape recorder.

### Measurements

The test words were sampled from a REVOX A77 stereo tape recorder through a low pass filter (based on Barr & Stroud filter models: 8 pole, Butterworth characteristic low pass filter) at a sampling rate of 10 kHz, and with cut-off frequencies at 5 kHz, into the Department of Linguistics Masscomp MC-500 computer. The samples were digitized and stored and analyzed using the ILS (Interactive Laboratory System) software package with the context set at 100, i.e., the analysis interval for extracting formant frequency data was set at 10 ms. After the API cepstrally-based pitch extraction analysis, the formant tracking was done on the spectrogram and formant values were calculated. In this study only the first and the second formants were considered. In deciding the formant value, the middle point of the clear formant structure was selected as a so called steady-state point in order to minimize the coarticulatory interference of the adjacent consonants. When there was an even number of frames for getting the vowel formant value, the average value of the two middle frames was taken.

### Statistics

Three-way ANOVAs with repeated measures were carried out using the statistical package BMDP. There are three independent variables with two levels each: pitch height (High, Low), vowel length (short, long) and syllable position (initial, final). The analyses were performed separately for the first and second formant for each vowel. Stepwise multiple regression analyses (BMDP 2R) were also performed to see how much of the total variance of the $F_1$ and $F_2$ may be accounted for by the variables considered in this experiment.

**The treatment of accent and pitch in this study**

In this experiment the effects of accent, pitch height, vowel length and syllable position on vowel variation of Japanese are studied. Vowel length and syllable position as independent variables seem to be straightforward enough. Accent and pitch, on the other hand, need some more explanation as there is a gap between the surface $F_0$ contour and the underlying abstract phonological representation, and the relationship between them seems to be quite complex.

In the present study with the nonsense words, the accent contrast was observed mostly in the initial syllable position: **ba'**ba and **ba**ba for the words with short vowels. For the words containing long vowels, only two types of intonational contours are possible, *ba'a baa* and *baa baa.* Thus, the accent contrast is observed only in the initial syllable position. Therefore, in order to assess the role of accent as a function of vowel quality, a separate set of ANOVAs were performed using the tokens in the initial syllable position.

Pitch height is far more complex. In assigning H and L to the syllables of the accented test words, H (accent H) was assigned to the accented syllable and L to the syllable following the accent (accent L). It should also be noted from Figure 6.2 that the accentual fall begins at the start of the stop closure of the second syllable in the word *ba'ba*, while it begins in the middle of the first long vowel in *ba'abaa*. The accentual fall seems to be aligned to the mora following the accented mora. For the unaccented words, L was assigned to the initial syllable (the interphrasal L%) and H was assigned to the second syllable (the phrasal H). For the finally accented short words, L was assigned to the first syllable (L%) and the accent H was assigned to the final syllable. Neither the difference between the accent H and the phrasal H, nor the difference between the accent L and the boundary L% have been considered in the statistical analyses. They are, despite

| | | speakers | | | | |
|---|---|---|---|---|---|---|
| | | KF | KO | JK | TK | YK |
| i | $F_1$ | 237 | 272 | 246 | 272 | 233 |
| | $F_2$ | 2347 | 2060 | 2357 | 1972 | 2178 |
| e | $F_1$ | 388 | 394 | 483 | 420 | 406 |
| | $F_2$ | 1968 | 1789 | 1915 | 1744 | 1914 |
| a | $F_1$ | 719 | 688 | 695 | 680 | 719 |
| | $F_2$ | 1119 | 1292 | 1174 | 1144 | 1163 |
| o | $F_1$ | 361 | 370 | 449 | 443 | 413 |
| | $F_2$ | 763 | 858 | 749 | 851 | 720 |
| u | $F_1$ | 296 | 306 | 343 | 312 | 279 |
| | $F_2$ | 1076 | 1243 | 1120 | 1228 | 1200 |

**Table 6.1.** The mean $F_1$ and $F_2$ values (in Hz) for the five vowels of Japanese (including long vowels) for each speaker.

their difference in phonological status and quality, treated in a binary term, either H or L.

### 6.2.2 Results

Figures 6.3 to 6.7 show the distribution of the five vowels for each speaker. No overlap in vowel types is observed. Though the context is limited to [bVbV] with the two V's being always the same in the carrier sentence *Kore wa ___ desu*, vowels show rather large variation. The midvowels /e/ and /o/, in particular, show a large spread in the first formant frequencies. Table 6.1 gives the mean $F_1$ and $F_2$ values for the five vowels of Japanese (including long vowels) for each speaker.

#### Accent

As explained at the beginning of this chapter, the only accent contrast observed across all the speakers is at the initial syllable position. For the vowels /e/, /a/, and /o/, in the initial syllable, the accented vowels have higher $F_1$ value than the unaccented vowels. Differences are about 30 to 40 Hz, and they are statistically

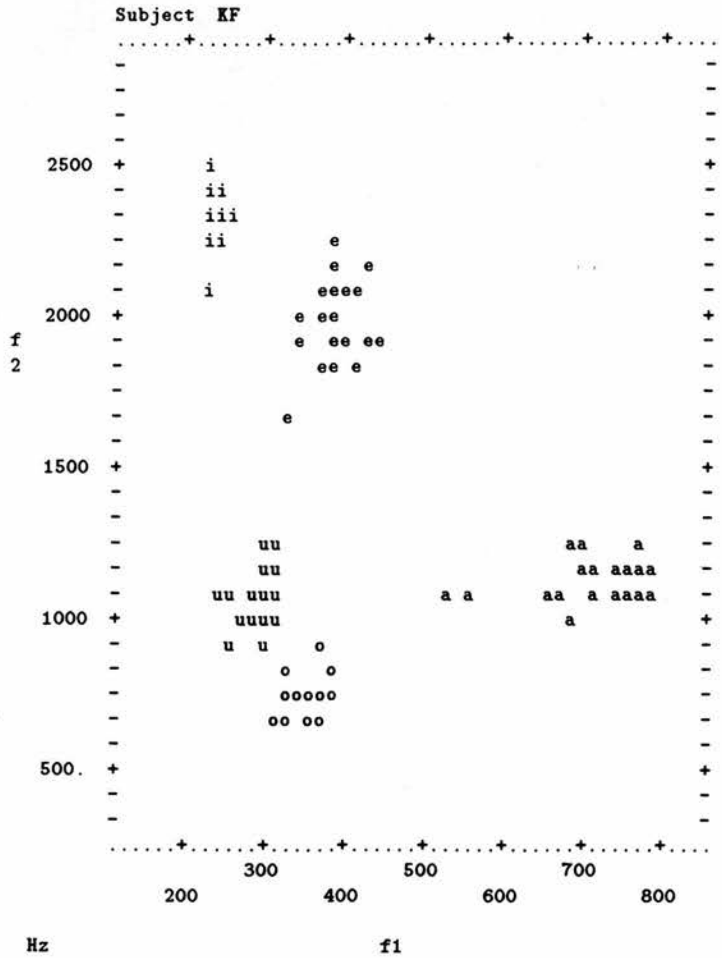



**Figure 6.3.** The scatterplot of the vowels /i, e, a, o, u/ produced by KF.

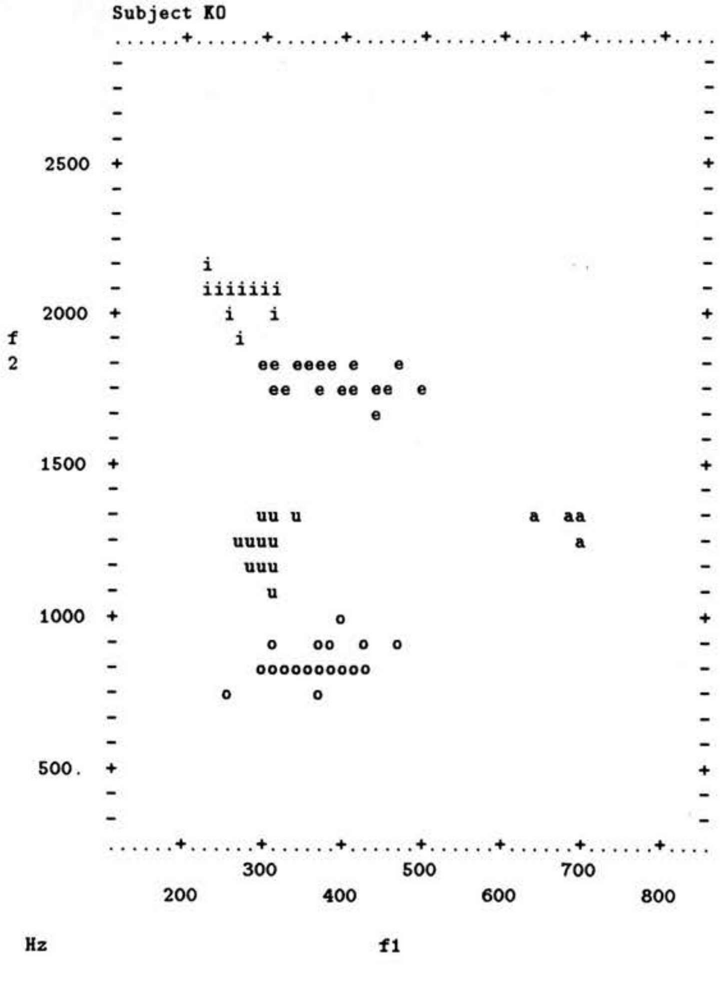



**Figure 6.4.** The scatterplot of the vowels /i, e, a, o, u/ produced by KO.

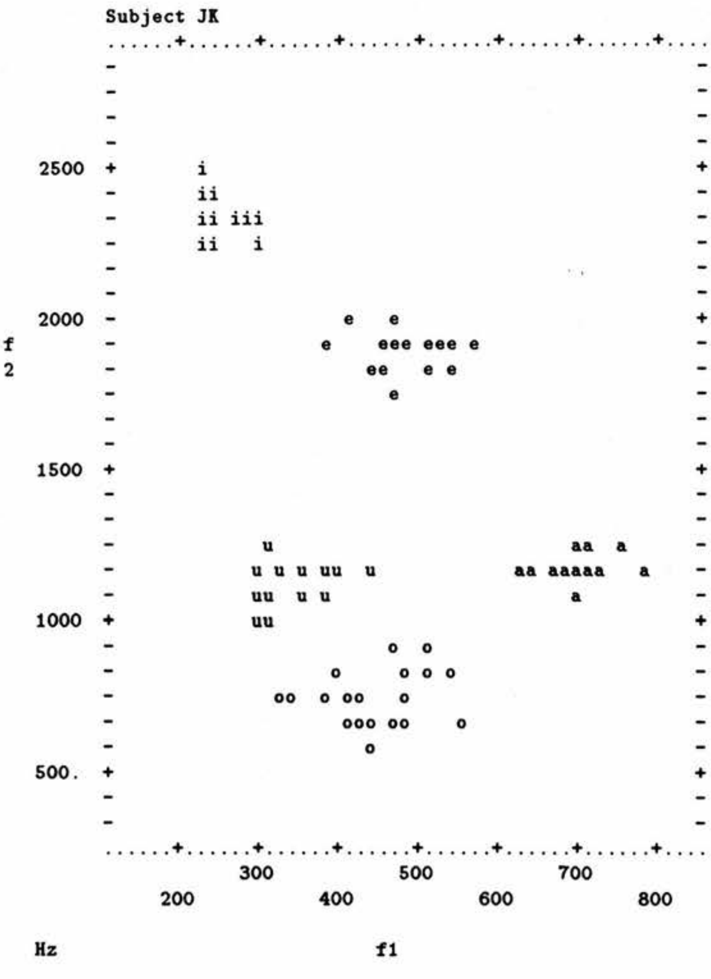



**Figure 6.5.** The scatterplot of the vowels /i, e, a, o, u/ produced by JK.

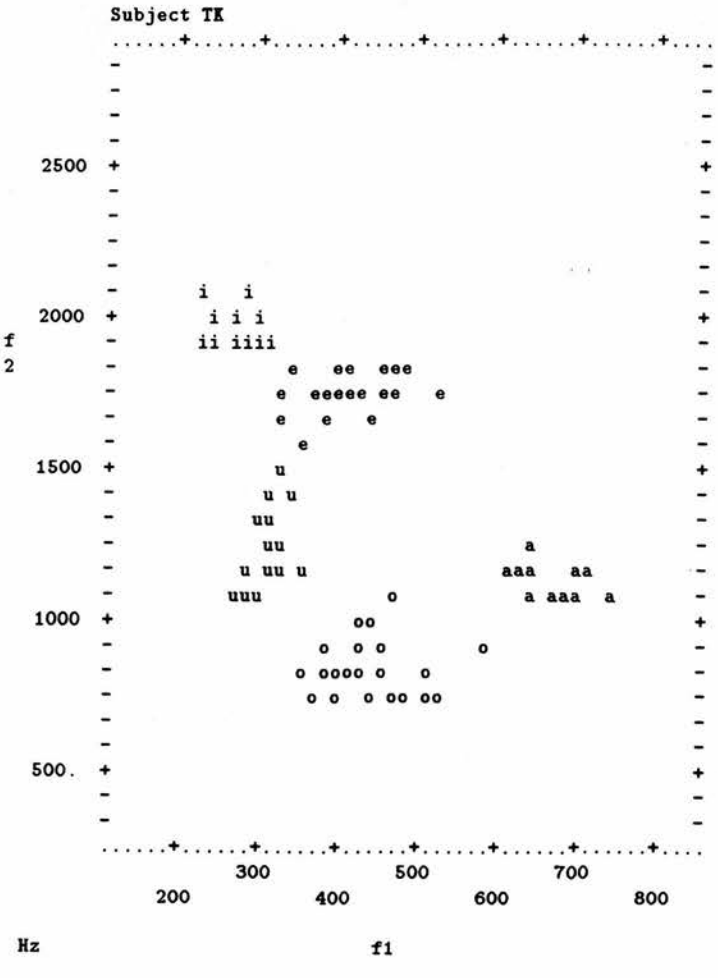



**Figure 6.6.** The scatterplot of the vowels /i, e, a, o, u/ produced by TK.

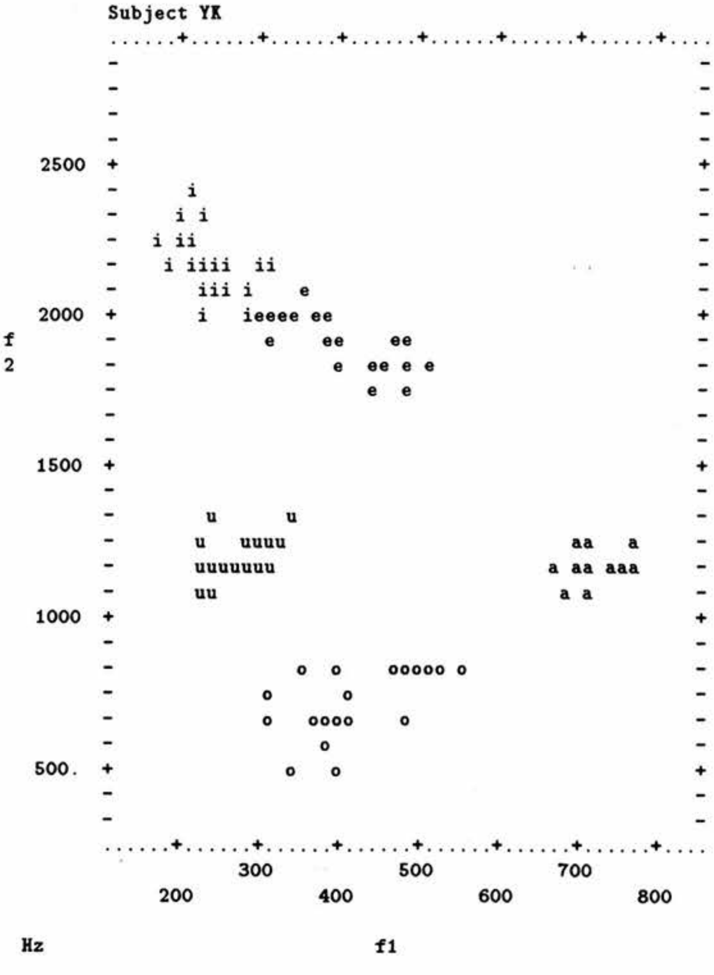



**Figure 6.7.** The scatterplot of the vowels /i, e, a, o, u/ produced by YK.

significant (/e/ $p < 0.01$, /a/ $p < 0.01$ and /o/ $p < 0.05$). The average $F_1$ and $F_2$ values for accented and unaccented syllables in the initial position for the five vowels across speakers are given in Table 6.2. For $F_2$, difference between accented and unaccented tokens is significant only for the vowel /i/.

| | | i | e* | a* | o* | u |
|---|---|---|---|---|---|---|
| $F_1$ | +ACCENT | 244 | 436 | 719 | 430 | 304 |
| | -ACCENT | 249 | 404 | 689 | 391 | 300 |
| | | i* | e | a | o | u |
| $F_2$ | +ACCENT | 2199 | 1878 | 1167 | 770 | 1147 |
| | -ACCENT | 2158 | 1855 | 1161 | 763 | 1162 |

**Table 6.2.** The mean $F_1$ and $F_2$ values (in Hz) for accented and unaccented initial syllable tokens for the five vowels across 5 speakers (N=30). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

When only the short tokens are considered, the trend observed above becomes clearer. Accented tokens have higher $F_1$ values than unaccented tokens for all of the five vowels. This may be due to a more open jaw position for accented tokens. However, the number of tokens is limited to three for each speaker. The difference reaches significant level for the vowels /e/ ($p < 0.007$), /a/ ($p < 0.001$), /o/ ($p < 0.01$), and /u/ ($p < 0.05$). The exception to this trend is observed for the vowel /i/ for JK and YK, for the vowel /o/ for KF and YK and for the vowel /u/ for TK and YK. Table 6.3 shows the average $F_1$ and $F_2$ values for the accented and

| | | i | e* | a* | o* | u* |
|---|---|---|---|---|---|---|
| $F_1$ | +ACCENT | 255 | 455 | 704 | 446 | 313 |
| | -ACCENT | 248 | 404 | 656 | 391 | 298 |
| | | i* | e | a | o | u |
| $F_2$ | +ACCENT | 2123 | 1828 | 1159 | 813 | 1151 |
| | -ACCENT | 2090 | 1805 | 1162 | 990 | 1155 |

**Table 6.3.** The mean $F_1$ and $F_2$ values (in Hz) for accented and unaccented initial syllable tokens for the short vowels across 5 speakers (N=15). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

unaccented initial tokens across speakers for the short vowels. Figure 6.8 shows the mean frequency values for the accented and unaccented tokens for the five short vowels across speakers. For $F_2$, the /i/ is again the only vowel that shows significant difference as a function of accent.

### Pitch Height

The main effect of pitch height for $F_1$ is significant for the front vowels /i/ ($p < 0.0184$), /e/ ($p < 0.0002$) and /a/ ($p < 0.0001$). For the two back vowels /o/ and /u/, the difference between High and Low pitched vowels was not significant. Table 6.4 shows the average values of the first and second formant frequencies for High and Low pitched tokens across five speakers for the five vowels. $F_1$ values are known to correlate with vowel height. Thus, $F_1$ values are higher for low vowels and lower for high vowels. Thus, when these vowels are centralized, it can be predicted that the two high vowels /i/ and /u/ will have higher $F_1$ value while the low vowel /a/ will have lower $F_1$ value. The distribution of the five vowels in the $F_1/F_2$ vowel space in Japanese suggests that the vowels /e/ and /o/ have rather low $F_1$ values. There is greater acoustic distance to the mid vowels /e/ and /o/ from the [+low] vowel /a/ than from the [+high] vowels /i/ and /u/ (see Figure 5.1). From this, it may be predicted that the vowels /e/ and /o/ will have higher $F_1$ when they are centralized. The results of the experiment show that the $F_1$ value is lower for the Low pitched tokens for the vowels /e/, /a/ and /o/, while it is higher for the Low pitched tokens for the two high vowels /i/ and /u/. The vowels /i/ and /a/ may be said to have centralized in Low pitched tokens.

The main effect of pitch for $F_2$ is significant only for the vowel /i/ ($p < 0.0021$). The second formant value correlates with the horizontal tongue position. The $F_2$ value is higher for the front vowels and lower for the back vowels. Thus, when the vowel is centralized, the $F_2$ value of the front vowel /i/ or /e/ is expected to lower while the $F_2$ value of the back vowel /o/ or /u/ is expected to rise. The $F_2$ value of the vowel /a/ may not vary largely. Because of the lip unrounding or compression, acoustically the $F_2$ value of /u/ is similar to that of /a/ in Japanese. However, the two vowels show different behaviour for centralization as the tongue body for the vowel /u/ is at the back and there is room for centralization while

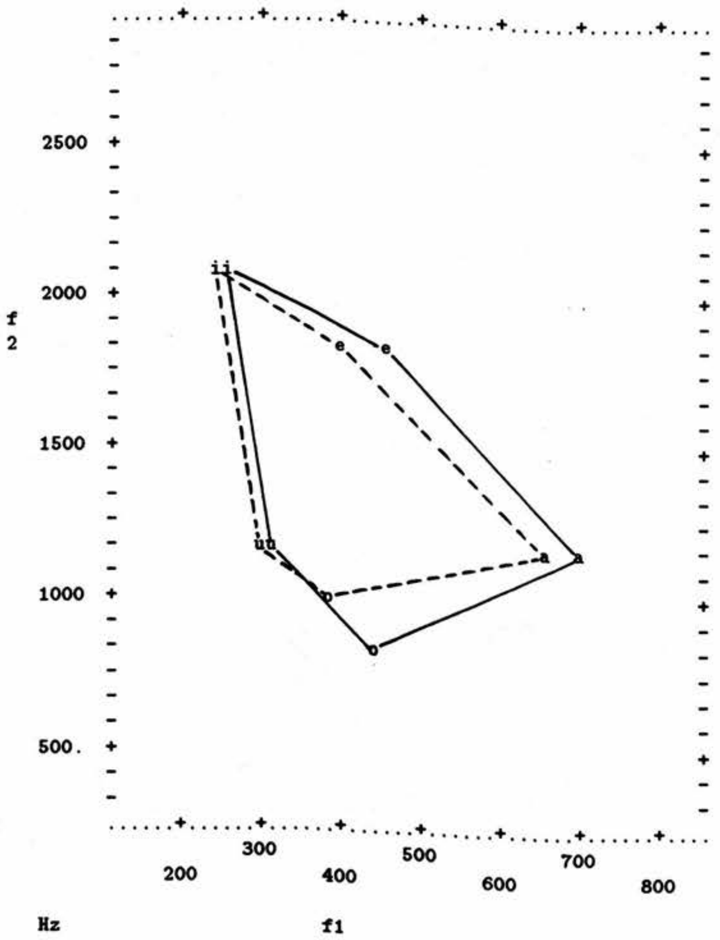



**Figure 6.8.** The mean $F_1$ and $F_2$ values plotted (in Hz) for the accented and unaccented short tokens across speakers. The solid line represents the accented tokens. The broken line represents the unaccented tokens.

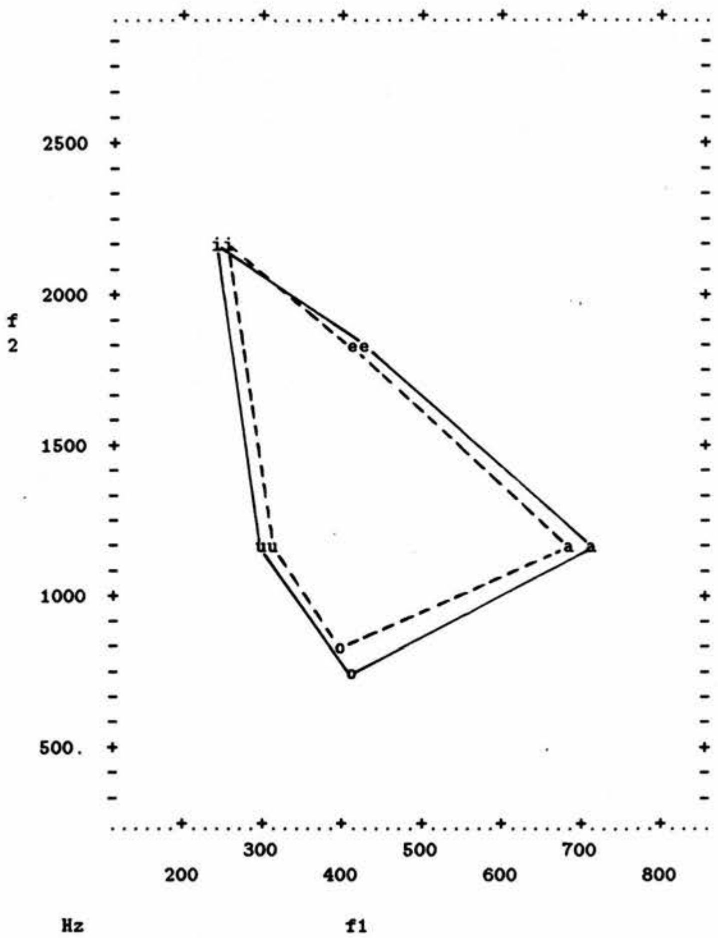



**Figure 6.9.** The mean $F_1$ and $F_2$ values plotted (in Hz) for the High pitched and Low pitched tokens across speakers. The solid line represents the High pitched tokens and the broken line represents the Low pitched tokens.

| | | i* | e* | a* | o | u |
|---|---|---|---|---|---|---|
| $F_1$ | HIGH | 246 | 428 | 709 | 411 | 306 |
| | LOW | 258 | 409 | 692 | 404 | 309 |
| | | i* | e | a | o | u |
| $F_2$ | HIGH | 2195 | 1872 | 1177 | 784 | 1165 |
| | LOW | 2169 | 1861 | 1179 | 792 | 1182 |

**Table 6.4.** The mean $F_1$ and $F_2$ values (in Hz) for High and Low pitched tokens for the five vowels across 5 speakers (N=60). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

the vowel /a/ is already central in the horizontal tongue dimension. The direction of change in $F_2$ value due to pitch height is, though not statistically significant for most of the vowels, towards centralization. However, Figure 6.9 shows that the difference in pitch is minimal.

### Length

The main effect of length for $F_1$ is significant for all of the five vowels /i/ ($p < 0.0064$), /e/ ($p < 0.0005$), /a/ ($p < 0.0000$), /o/ ($p < 0.0005$) and /u/ ($p < 0.0001$). The low vowel /a/ has lower $F_1$ value when it is short while the other four vowels have higher $F_1$ value when they are short. The short vowels may be characterized as more central than the long vowels in $F_1$. The average first and second formant frequencies for long and short vowels across five speakers for the five vowels are given in Table 6.5. The subject TK has a different trend from the others in that he has slightly lower $F_1$ value for the short tokens even for the high vowels /i/ and /u/.

The main effect of vowel length for $F_2$ is significant for the vowels /i/ ($p < 0.0000$), /e/ ($p < 0.0000$), /o/ ($p < 0.0000$) and /u/ ($p < 0.0016$). The difference is about 40Hz for the vowel /u/ to 100Hz for the other three vowels. In $F_2$ as well, the short vowels may be characterized as more central than the long vowels. Figure 6.10 shows that the short vowels tend to cluster at the inner part of the acoustic vowel space though some overlap between short and long tokens is observed. The mean frequency values for the long and short vowels across

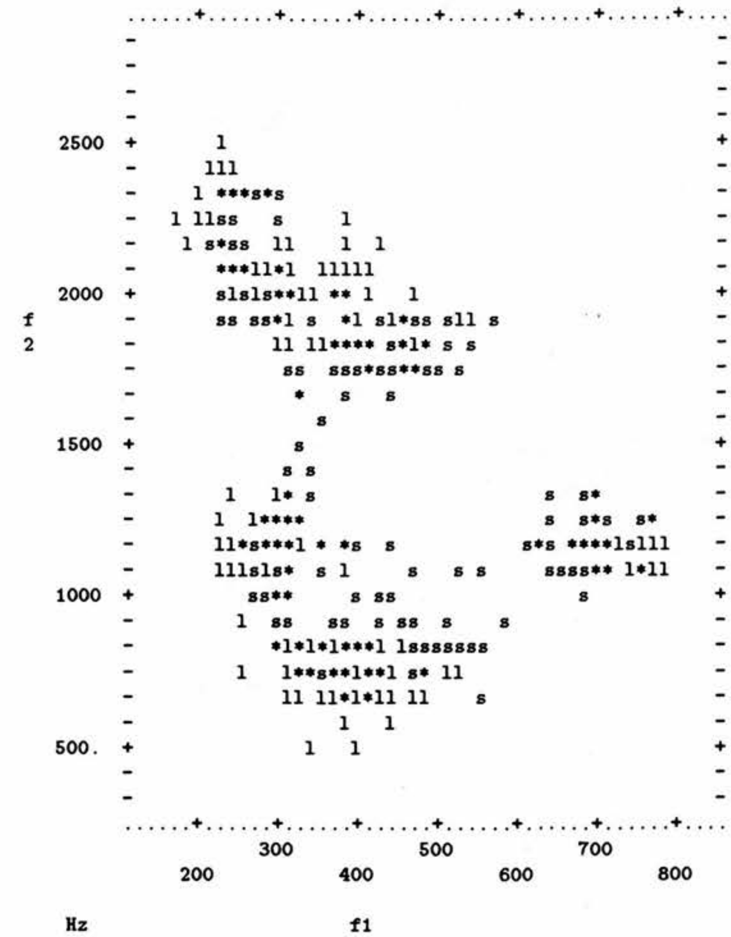



**Figure 6.10.** The scatterplot of the long and short tokens across speakers. Symbols l and s represent long and short vowels respectively. Symbol * means that tokens from the two groups showed overlap.

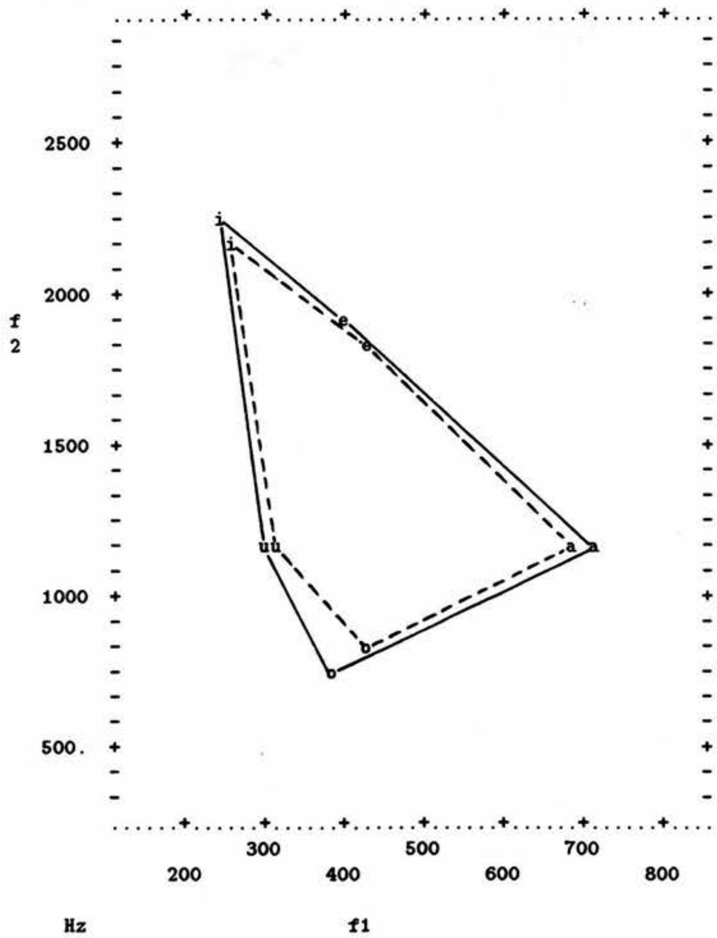



**Figure 6.11.** The mean $F_1$ and $F_2$ values plotted (in Hz) for the long and short tokens across speakers. The solid line represents the long tokens and the broken line represents the short tokens.

| | | i* | e* | a* | o* | u* |
|---|---|---|---|---|---|---|
| $F_1$ | LONG | 248 | 405 | 717 | 390 | 298 |
| | SHORT | 258 | 431 | 683 | 424 | 316 |
| | | i* | e* | a | o* | u* |
| $F_2$ | LONG | 2233 | 1918 | 1173 | 740 | 1152 |
| | SHORT | 2132 | 1815 | 1183 | 836 | 1194 |

**Table 6.5.** The mean $F_1$ and $F_2$ values (in Hz) for long and short tokens for the five vowels across 5 speakers (N=60). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

| | | i* | e | a | o | u* |
|---|---|---|---|---|---|---|
| $F_1$ | INITIAL | 246 | 420 | 704 | 411 | 302 |
| | FINAL | 258 | 417 | 697 | 403 | 313 |
| | | i | e | a* | o* | u* |
| $F_2$ | INITIAL | 2176 | 1866 | 1164 | 767 | 1154 |
| | FINAL | 2188 | 1867 | 1193 | 809 | 1192 |

**Table 6.6.** The mean $F_1$ and $F_2$ values (in Hz) for tokens in the initial and final syllable for the five vowels across 5 speakers (N=60). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

speakers are plotted in Figure 6.11.

### Syllable Position

The main effect of syllable position for $F_1$ is significant for the two high vowels /i/ ($p < 0.01$) and /u/ ($p < 0.0007$). The vowels /i/ and /u/ have higher $F_1$ value when they are in the final syllable. The average $F_1$ and $F_2$ values for the tokens in the initial and final syllables across five speakers for the five vowels are given in Table 6.6. The vowels /e/, /a/ and /o/ have lower $F_1$ values in the final syllable. The main effect of syllable position for $F_2$ is significant for the vowels /a/ ($p < 0.0000$), /o/ ($p < 0.0003$) and /u/ ($p < 0.0003$). The $F_2$ value is higher in the final syllable for all of the five vowels. The effect of syllable position on the Japanese vowels is very small (Figure 6.12).

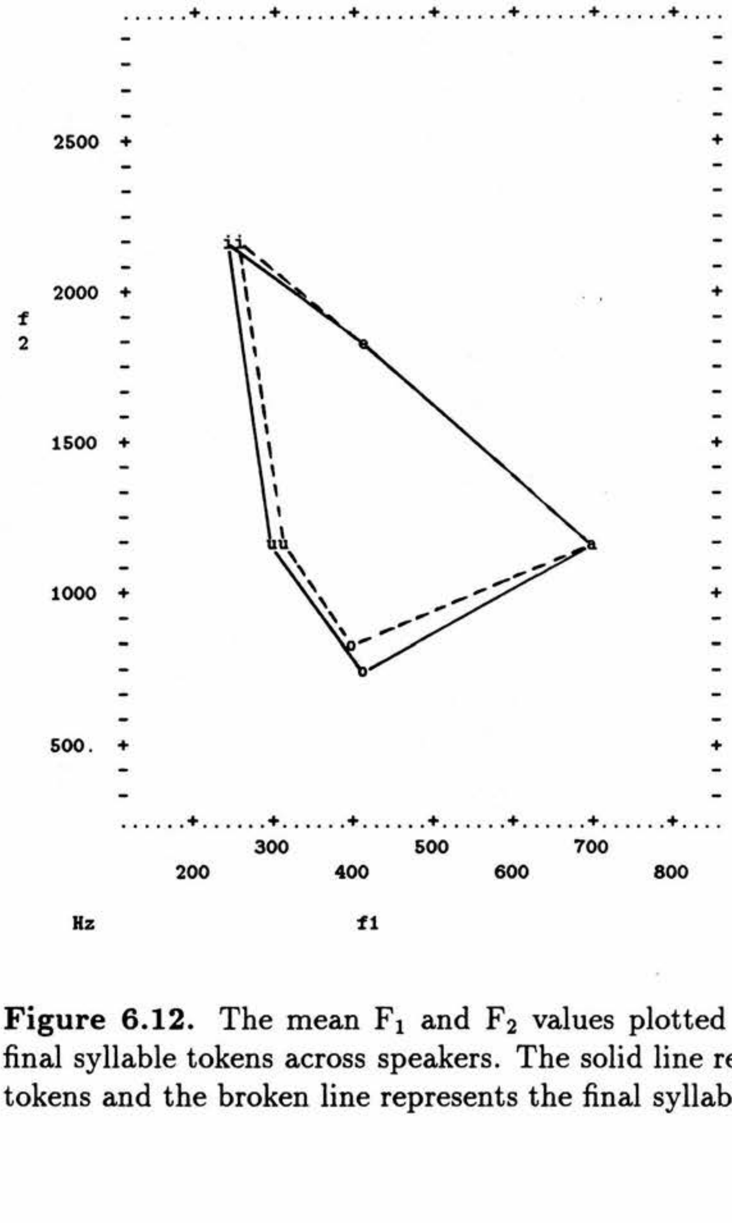



**Figure 6.12.** The mean $F_1$ and $F_2$ values plotted (in Hz) for the initial and final syllable tokens across speakers. The solid line represents the initial syllable tokens and the broken line represents the final syllable tokens.

| | | i | e | a* | o | u |
|---|---|---|---|---|---|---|
| $F_1$ | LIH | 197 | 328 | 723 | 370 | 230 |
| | SFL | 242 | 431 | 685 | 477 | 307 |
| | | i | e | a | o* | u |
| $F_2$ | LIH | 2329 | 2014 | 1195 | 558 | 1119 |
| | SFL | 2125 | 1857 | 1153 | 780 | 1191 |

**Table 6.7.** The mean $F_1$ and $F_2$ values (in Hz) for LIH and SFL tokens for the five vowels across 5 speakers (N=15). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

### Speakers

The main effect of speaker was significant for both first and second formant frequencies for all of the five vowels ($p < 0.0005$ for $F_1$, and $p < 0.0001$ for $F_2$) when each vowel was considered separately. However, when the data for the five vowels were considered together, there was no significant effect of speaker for either $F_1$ ($p = 0.9515$) or $F_2$ ($p = 0.9959$). For $F_1$, the interaction between speaker and the three variables, pitch ($p < 0.0001$), length ($p < 0.0001$) and syllable position ($p < 0.0543$) were significant. That is, there was a considerable inter-speaker variability in $F_1$ variance as a function of the three variables. For $F_2$, the interaction between speaker and syllable position ($p < 0.0046$) was significant, but the interaction between speaker and pitch ($p = 0.9710$), or speaker and length ($p = 0.6849$) was not significant. That is, the effects of pitch were similarly small for all of the five speakers while the effects of vowel length were significant for all the speakers.

### Combined Effect

When the three independent variables, pitch, length and syllable position are combined, a long, initial, High pitched vowel (LIH) is more peripheral than a short, final, Low pitched vowel (SFL) as shown in Figure 6.13. In this case pitch height corresponds to accent as well. This trend is observed for all of the five vowels for all the five speakers. Differences are statistically significant only for the vowel /a/ for $F_1$, and for the vowel /o/ for $F_2$ (the post hoc scheffe test: p <

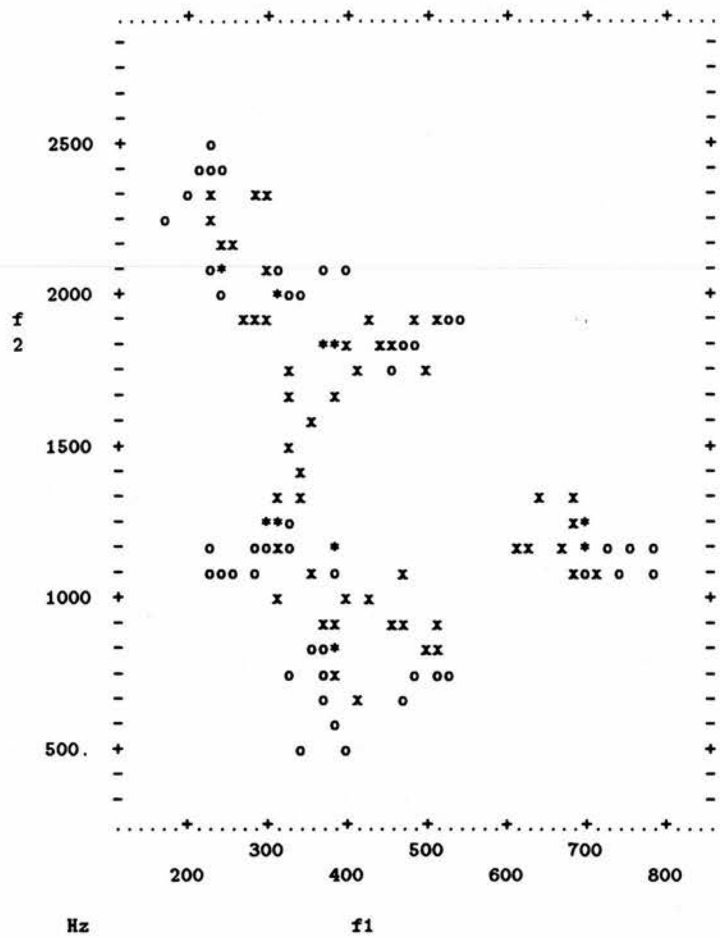



**Figure 6.13.** The mean $F_1$ and $F_2$ values plotted (in Hz) for long, intitial, High pitched (LIH, symbol = o) and short, final, Low pitched (SFL, symbol = x) tokens across speakers. The symbol * means that tokens form the two groups showed overlap.

| | $F_1$ | | | $F_2$ | | |
|---|---|---|---|---|---|---|
| | variable entered | $r^2$ | change in $r^2$ | variable entered | $r^2$ | change in $r^2$ |
| i | pitch | 0.0374 | | speaker | 0.1247 | |
| | syllable | 0.0707 | 0.0332 | length | 0.2088 | 0.0841 |
| e | accent | 0.0572 | | length | 0.1973 | |
| | length | 0.0912 | 0.0340 | speaker | 0.2306 | 0.0333 |
| a | length | 0.1577 | | syllable | 0.0397 | |
| | accent | 0.2903 | 0.1327 | | | |
| o | speaker | 0.1550 | | length | 0.2751 | |
| | accent | 0.2433 | 0.0882 | syllable | 0.3300 | 0.0548 |
| | length | 0.2992 | 0.0559 | | | |
| | pitch | 0.3262 | 0.0270 | | | |
| | syllable | 0.3493 | 0.0231 | | | |
| u | length | 0.0636 | | speaker | 0.1002 | |
| | | | | length | 0.1417 | 0.0414 |
| | | | | syllable | 0.1741 | 0.0325 |

**Table 6.8.** A summary table for the results of stepwise multiple regression analyses (BMDP 2R). The dependent variables are the midpoint $F_1$ and $F_2$ values for the vowels /i, e, a, o, u/. The independent variables are accent, pitch, vowel length, syllable position and speaker.

0.05). It seems that the three variables affect vowels accumulatively. Table 6.7 show the mean $F_1$ and $F_2$ values for the LIH and SFL tokens across speakers.

### Multiple Regression Analyses

Stepwise multiple regression analyses were performed with the vowel midpoint $F_1$ and $F_2$ values as dependent variables, and accent, pitch, length, syllable position and speaker as independent variables. In stepwise analyses, variables are entered in the order of the importance in predicting the variance of the dependent variable. Variables are not entered into the equation when they do not significantly increment the predictability. The results are summarized in Table 6.8. All the five variables combined together account for 35% of the total variance of the vowel /o/'s $F_1$ values. This is the highest $r^2$ value in the table. Next to this, length and syllable position account for 33% of the total variance of the $F_2$ values

| | | | | | | |
|---|---|---|---|---|---|---|
| | | i | e* | a | o* | u* |
| HIGH | LONG | 239 | 405 | 717 | 391 | 293 |
| | SHORT | 253 | 450 | 701 | 430 | 318 |
| | | i | e | a* | o* | u |
| LOW | LONG | 258 | 406 | 618 | 389 | 303 |
| | SHORT | 259 | 412 | 666 | 418 | 314 |

**Table 6.9.** The mean $F_1$ values (in Hz) for the long and short vowels for the High and Low pitched tokens across 5 speakers (N=30). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

of the vowel /o/. The vowel /o/ seems to be most readily affected by the variables considered in this experiment. A combination of length and speaker seems to be the best predictor of the $F_2$ values of the vowels /i/, /e/ and /a/. Syllable position significantly increases the predictability of $F_2$ values for the back vowels /o/ and /u/. Neither accent nor pitch accounts for $F_2$ variance. For $F_1$, accent and length seem to be the two important predictors of the formant frequency variability.

### Interactions

**$F_1$: pitch height × vowel length** The interaction between pitch height and vowel length is significant for the vowels /i/ ($p < 0.0364$), /e/ ($p < 0.0181$) and /a/ ($p < 0.0011$). For both High and Low pitched tokens, the tendency seems to be that of centralization for short vowels except for the vowel /a/. The vowel /a/ behaved differently in the two pitch conditions. The trend for centralization was more pronounced for the High pitched tokens than for Low pitched ones (Table 6.9).

**$F_1$: pitch height × syllable position** For $F_1$, the interaction between pitch height and syllable position is significant for all of the vowels /i/ ($p < 0.0067$), /e/ ($p < 0.0362$), /a/ ($p < 0.0012$), /o/ ($p < 0.0015$) and /u/ ($p < 0.0221$). For the High pitched tokens the contrast of the initial and final token is the contrast of the accent H and phrasal H (for KF and JK, the accent H). For the Low

| | | i | e | a* | o* | u |
|---|---|---|---|---|---|---|
| HIGH | INITIAL | 244 | 436 | 719 | 430 | 304 |
| | FINAL | 248 | 420 | 700 | 391 | 308 |
| | | i* | e | a | o | u* |
| LOW | INITIAL | 249 | 404 | 689 | 391 | 300 |
| | FINAL | 268 | 414 | 694 | 416 | 318 |

**Table 6.10.** The mean $F_1$ values (in Hz) for tokens in the initial and final syllable for the High and Low pitched vowels across 5 speakers (N=30). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

| | | i | e | a* | o* | u |
|---|---|---|---|---|---|---|
| LONG | INITIAL | 241 | 410 | 727 | 404 | 298 |
| | FINAL | 255 | 401 | 707 | 377 | 298 |
| | | i | e | a | o | u* |
| SHORT | INITIAL | 252 | 430 | 680 | 418 | 305 |
| | FINAL | 260 | 433 | 687 | 430 | 327 |

**Table 6.11.** The mean $F_1$ values (in Hz) for tokens in the initial and final syllable for the long and short vowels across 5 speakers (N=30). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

pitched tokens, the contrast of the initial and final token is the contrast of the boundary L% and the accent L. The high vowels /i/ and /u/ have higher $F_1$ value in the final syllable regardless of the pitch condition. For Low pitched tokens, the $F_1$ values are higher in the final syllable for all of the five vowels (Table 6.10). This may suggest that the accent L tokens have higher $F_1$ than the boundary L% tokens. The difference is significant for the vowels /i/ and /u/ ($p < 0.01$). The higher $F_1$ may be a correlate of accent (the accent L in this case). Accented tokens may be characterized by a more open articulatory configuration regardless of their pitch. That is, the HL, the accent as a whole, may be realized with a more open jaw position.

**$F_1$: vowel length × syllable position** The interaction between vowel length and syllable position is significant for the vowels /a/ ($p < 0.0035$), /o/ ($p < 0.0005$) and /u/ ($p < 0.005$). All of the five vowels have higher $F_1$ value in the final syllable when they are short (Table 6.11). The difference is significant only for the vowel /u/ ($p < 0.01$). For short tokens, there are HL (accented) tokens and LH (accented and unaccented) tokens. All the final L tokens are the accent L tokens. Besides, there are realizations of both the accent H (KF and JK) and the phrasal H (KO, TK, YK) in the final syllable for the LH tokens. It is likely that these finally accented H tokens are biasing the results towards higher $F_1$ values in the final syllable. When these tokens were eliminated from the analyses, it was shown that only the vowels /i/ and /u/ had higher $F_1$ values in the final syllable. In summary, the first formant value is higher in the final syllable when the vowel is short or Low pitched.

**$F_2$: pitch height × vowel length** There is no interaction between pitch height and vowel length. The short vowels are more central than the long vowels regardless of pitch condition in $F_2$. The long vowels and short vowels are significantly separated, and the difference is significant for both High and Low pitched tokens ($p < 0.05$ and $p < 0.01$ respectively).

**$F_2$: pitch height × syllable position** The $F_2$ is higher when the vowel is in the final syllable for the back vowels /a/, /o/ and /u/ regardless of pitch

| | | i | e | a | o* | u |
|---|---|---|---|---|---|---|
| HIGH | INITIAL | 2199 | 1878 | 1166 | 770 | 1147 |
| | FINAL | 2190 | 1867 | 1187 | 798 | 1183 |
| | | i | e | a* | o* | u |
| LOW | INITIAL | 2154 | 1855 | 1161 | 763 | 1162 |
| | FINAL | 2185 | 1867 | 1198 | 820 | 1201 |

**Table 6.12.** The mean $F_2$ values (in Hz) for tokens in the initial and final syllable for the High and Low pitched vowels across 5 speakers (N=30). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

condition. For the Low pitched vowels the $F_2$ is higher in the final syllable for all of the five vowels. Statistically the difference is significant only for the vowels /a/ ($p < 0.05$) and /o/ ($p < 0.01$). When the vowels are High pitched, the $F_2$ value is lower in the final syllable for the front vowels /i/ and /e/ and higher for the back vowels /a/, /o/ and /u/. This is more in accordance with the centralization tendency. The $F_2$ values for the initial and final syllables are given separately for the High and Low pitched tokens in Table 6.12.

**$F_2$: vowel length $\times$ syllable position** For $F_2$, the interaction between syllable position and vowel length is significant for the vowels /i/ ($p < 0.0000$), /a/ ($p < 0.0395$), /o/ ($p < 0.0032$) and /u/ ($p < 0.0013$). For the short vowels, $F_2$ is higher in the final syllable except for the vowel /e/. And the difference due to syllable position is significant for the other four vowels /i/ ($p < 0.01$), /a/ ($p < 0.05$), /o/ ($p < 0.01$) and /u/ ($p < 0.01$). For the long vowels, the difference due to syllable position is not significant. The $F_2$ values for the initial and final syllables are given separately for the long and short vowels in Table 6.13. $F_1$ as well as $F_2$ seems to rise in the final syllable when the vowel is short or Low pitched. The only exception to this is observed for the $F_2$ value of the short vowel /e/ where the $F_2$ value is slightly higher in the initial syllable (1817Hz) than in the final syllable (1813Hz).

| | | i | e | a | o | u |
|---|---|---|---|---|---|---|
| LONG | INITIAL | 2250 | 1916 | 1167 | 732 | 1156 |
| | FINAL | 2215 | 1920 | 1179 | 749 | 1148 |
| | | i* | e | a* | o* | u* |
| SHORT | INITIAL | 2102 | 1817 | 1161 | 801 | 1153 |
| | FINAL | 2161 | 1813 | 1206 | 870 | 1236 |

**Table 6.13.** The mean $F_2$ values (in Hz) for tokens in the initial and final syllable for the short and long vowels across 5 speakers (N=30). The symbol * shows that the difference has reached the significant level of $p < 0.05$.

### Summary of results

It should be noted that vowels are extremely variable in nature and under this background noise variability, it is difficult to see that vowels under one condition are significantly different from those under another condition, specially when the three variables in the experiment seem to be interacting in a complex manner. It should also be remembered that speakers did not always act uniformly.

Each vowel in the vowel inventory of Japanese /i, e, a, o, u/ also behaved differently as a function of the variables considered here. Table 6.14 is a summary table for the significance of the variables considered here for each vowel. Vowel length was the only variable which affected all the vowels in this experiment. So far a linear scale in frequency Hz has been used for statistical analyses. However, it is well known that human perception does not operate linearly. At a lower frequency range, people can detect a smaller change in frequency Hz than at a higher frequency range. A mel scale is designed to approximate this pattern of human perception.[2] Its units are equal intervals of perceived pitch rather than frequency. ANOVAs were rerun using a mel scale for all the variables listed in Table 6.14. The results obtained by using a mel scale are given in the parentheses when they were different from those obtained by the linearly scaled data. In general there was no major difference in the trend observed. By using a mel scale, however, the effect of accent on $F_2$ became negligible.

[2]The equation for converting frequency Hz into a mel scale is $mel = 2595 \times log(1 + (f/700))$ where $f$ is the formant frequency in Hz (Makhoul & Cosell 1976).

| | | Accent short V's N = 15 | Pitch N = 60 | Length N = 60 | Syllable N = 60 | Combined N = 15 |
|---|---|---|---|---|---|---|
| $F_1$ | i | − | + | + | + | + |
| | e | + | + | + | − | − |
| | a | + | + | + | − | + |
| | o | + | − | + | − | − |
| | u | +(−) | − | + | + | + |
| $F_2$ | i | +(−) | + | + | − | +(−) |
| | e | − | − | + | − | + |
| | a | − | − | −* | + | − |
| | o | − | − | + | + | + |
| | u | − | − | + | + | −(+) |

**Table 6.14.** A summary table for the significance of the variables considered here for each vowel. The symbol '+' means 'significant'. The symbol '−' means 'not significant'. *The vowel /a/ is not significantly affected by the vowel length as it is already a central vowel in $F_2$. The signs within ( ) show the significance of the variable concerned when a mel scale was used instead of a linear scale.

1. Significant difference in $F_1$ due to the presence of accent was observed for the vowels /e/, /a/ and /o/ (and /u/ when only the short vowels are considered). It seems that when a vowel is accented, the $F_1$ value is raised. Accented vowels seem to be characterized by a more open articulation. No significant effect of accent was observed on $F_2$. The $F_2$ values seem to remain unaffected by accent.

2. The effect of pitch height on $F_1$ is significant for the front vowels /i/, /e/ and /a/. However, the subjects behaved differently except for the vowel /a/, which was lower in $F_1$ for Low pitched syllables for all of the subjects. The effect of pitch height on $F_2$ is significant for the vowel /i/ only. The vowel /i/ had a lower $F_2$ value when it was Low pitched. The speakers again showed different behaviour except for the vowels /i/ and /a/. In general the effect of pitch height on vowel formant values seems to be very small.

3. The effect of vowel length on both $F_1$ and $F_2$ is significant for all the vowels except for the vowel /a/ in $F_2$. The short vowels tend to occupy more central region of the acoustic vowel space than the long vowels, and this behaviour was uniformly observed across speakers.

4. The effect of syllable position was significant for the vowels /i/ and /u/ for $F_1$, and for the vowels /a/, /o/ and /u/ for $F_2$. There was a considerable amount of variability in behaviour among the speakers. It seems that for both $F_1$ and $F_2$, vowels in the final syllable have higher formant value when they are either Low pitched or short.[3] Higher $F_1$ values in the final syllable for Low pitched or short tokens may be attributed to accent effect.

5. When the effects of pitch height, length and syllable position are combined, short, Low pitched (unaccented) tokens in the final syllable occupy more central part of the acoustic vowel space than long, High pitched (accented) tokens in the initial syllable. The three effects seem to be accumulative.

[3]It is not quite clear why $F_2$ values were higher in the final syllable for Low pitched and short tokens.

### 6.2.3 Discussion

As discussed in the previous chapter (see 5.4), Beckman (1986) suggests that pitch is the sole correlate of accent in Japanese. Unlike English, variables such as duration, amplitude and vowel quality, seem to play a minimal role in cueing accent. From this, it may be inferred that accent in Japanese may not affect vowel quality in a linguistically significant way. Indeed, the results of the present experiment suggest that accent has minimal effects on vowel quality, particularly in $F_2$. For $F_1$, it was observed that accented vowels have higher frequency values than unaccented vowels. Differences were significant for the vowels /e, a, o, u/ (when only short vowels were considered) and these differences ranged from 15 Hz for /e/ to around 50 Hz for the other three vowels across five speakers.[4] Accented vowels may have a generally more open articulation. Pitch also seems to correlate very little with vowel quality.[5] Where significant effects of pitch on vowel formant frequencies were observed, differences as a function of pitch were minimal: for $F_1$, 12 Hz, 19 Hz and 17 Hz for /i, e, a/ and for $F_2$, 26 Hz for /i/.

On the other hand, vowel duration seems to have significant effects on vowel quality though traditionally it has been assumed that phonemically distinctive short and long vowels of Japanese are similar in quality. A spectral pattern taken at the middle of a single /o/ in the phrase *so***t*o***dewa* (outside) shows that it is central in quality, i.e., the three peaks are more or less evenly spaced. On the other hand, a spectral pattern taken at the middle of two /o/'s across a word boundary in *kasakasat***o** **o***to* (sound of rustling) shows a more /o/-like spectral pattern with the first and the second peaks close together (Figure 6.14). Keating & Huffman (1984) have also observed that the /oo/ tokens were more peripheral in the acoustic vowel space than /o/ tokens.[6] When a segmented single /o/

[4]In this discussion, frequency values are expressed in a linear scale.

[5]However, a prosodic category such as an accentual phrase which is defined by a pitch contour may constrain the extent of coarticulation in Japanese. Jun (1993) has observed that an accentual phrase is the domain of voicing assimilation in Korean. It was also suggested that a stress foot may be the domain of coarticulation in English (Fowler 1981). Fowler's hypothesis states that segments other than stressed vowels may be subsumed within the domain of production of a stressed vowel. It is plausible that prosodic categories which seem to involve some sort of planning in speech production may be the domain of coarticulation in languages.

[6]In their data corpus, the long vowels consisted almost exclusively of /oo/ and /uu/. The /uu/ tokens occurred mainly following a palatal consonant, and as a result had high $F_2$ values

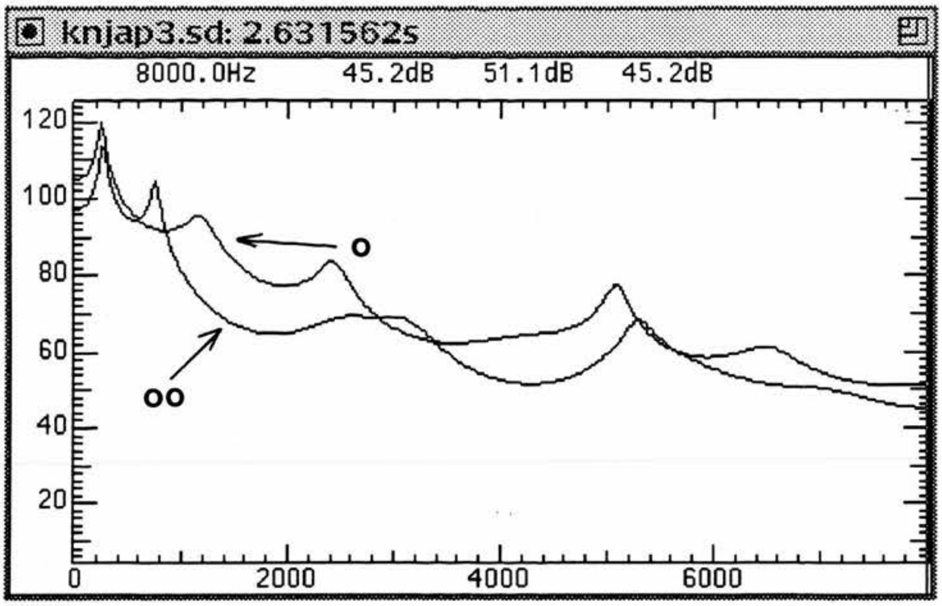



**Figure 6.14.** A spectral pattern taken at the middle of a single /o/ in the word *sotodewa* (the top one) and that taken at the middle of two /o/'s across a word boundary in *kasakasato oto.*

was listened to, it often gave an impression of /u/ or at times of a schwa-like sound, whereas a double /oo/ was heard as /o/. These effects may be interpreted either as centralization or increased contextual effect for $F_2$. The results obtained by Lindblom (1963:Fig. 2:1775) support the increased contextual assimilation hypothesis. (See Figure 3.1.) In either case, 'target undershoot' is what seems to be happening when the vowel length is shorter.

It was observed that the tendency for centralization became more prominent when the effects of the three variables, accent (pitch), length and syllable position, were combined. This was partly supported by the results of the stepwise multiple regression analyses.

Here the issue of centralization or contextual assimilation must be reconsidered. Delattre (1969) studied vowel reduction of the four languages, English, German, French and Spanish. He used minimal pairs such as, *competing vs.*

filling in more of the high central part of the formant space.

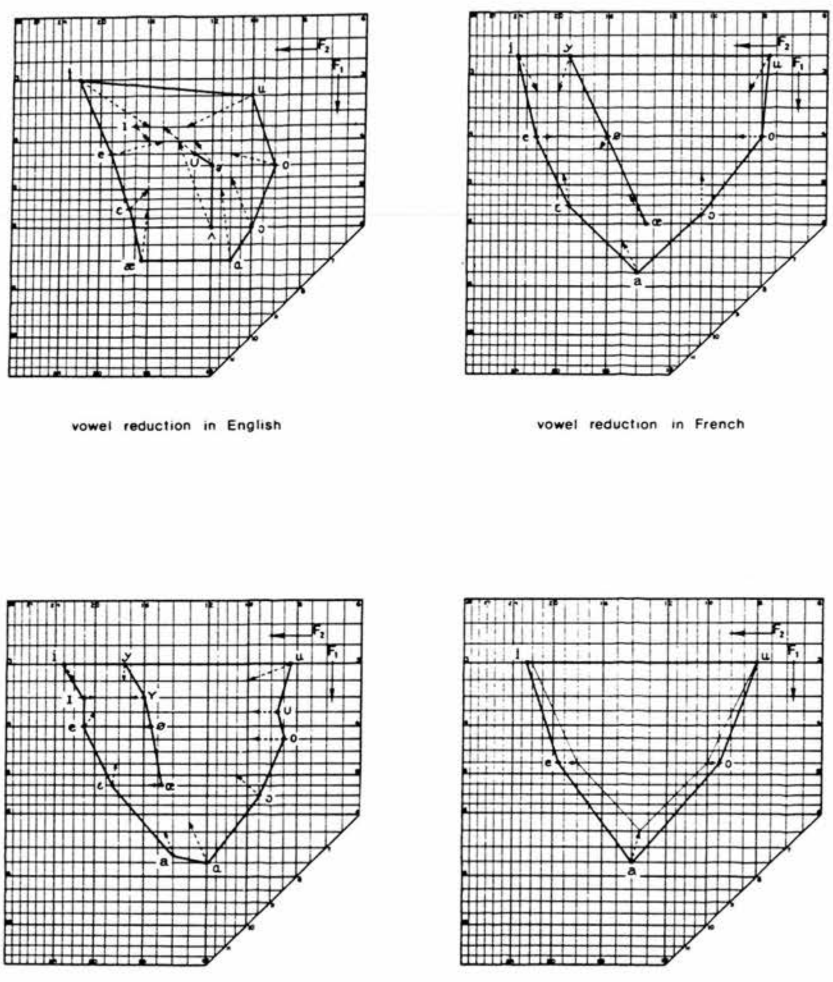



Figure 1. A comparison of the acoustic centering of unstressed vowels in four languages. The greater length of broken lines in English indicates a higher degree of vowel reduction in this language than in the three others.

**Figure 6.15.** Figure taken from Delattre (1969:p309).

*competition* in the four languages. Figure 6.15 taken from Delattre (1969:p309) illustrates the results of his study. Each language seems to have its own degree and pattern of vowel reduction. English shows the greatest vowel reduction while Spanish shows the least. What we observe here is centralization for most vowels and fronting for some central vowels, e.g., /$\phi$, œ/ for German and French. The pattern observed for Spanish is closest to the pattern observed for Japanese in the present experiment, the high vowels /i/ and /u/ showing little effect of accent.

If vowel reduction is contextual assimilation rather than a natural tendency for vowels to degenerate into the central quality of the rest position, as suggested by Lindblom (1963), Nord (1974), Bergem (1993, 1994) and the present study, then centralization may be considered as a measure of the extent of contextual assimilation. Then Figure 6.15 illustrates the extent to which accent affects contextual assimilation in the four languages. Following this line of thought, the results of the present experiment suggest that accent does not affect the extent of contextual assimilation in the $F_2$ values of the Japanese vowels while accent may be correlated with the openness of articulation.

## 6.3 Conclusion

The vowel variation due to the factors considered in this experiment is small. Among the three factors, the vowel length seems to have the greatest effect on vowel formant frequencies. When the three factors are added up, there seems to emerge the tendency towards centralization for the vowels which are short, in the final syllable and Low pitched (or unaccented). It was also noted that for the short vowels, where accent contrast was observed, $F_1$ value was higher for the vowels on which accent falls, suggesting a more open articulation for accented vowels. However, in general the effects of these factors are subsumed under the noise variability and the complex factorial interactions. Furthermore, in many cases the speakers showed variability in behaviour. The effect of the factors considered here seems to vary from speaker to speaker. They are highly idiosyncratic though collectively a certain pattern was observed and the pattern suggested centralization, though minimal, for short, Low pitched vowels in the final syllable.

# Chapter 7

# V-to-V Coarticulation in Japanese

## 7.1 Introduction

The V-to-V coarticulation in Japanese is explored in the present chapter. There are three main issues addressed in this chapter. The first one is the effect of accent on the magnitude of V-to-V coarticulation. The second one is the temporal effect and the direction of the V-to-V coarticulation across a segment. Thirdly, the effect of secondary articulation, e.g., palatalization, in blocking the V-to-V coarticulation is studied.

In the VCV articulation, the nature of the middle consonant and the nature of the affected vowel ($V_1$ for anticipatory and $V_2$ for carryover effects) seem to determine the extent to which the effects of the transconsonantal vowel may penetrate through the affected segment. The effects would be more pronounced when both of them are transparent in nature. Also, rhythmic constraints, i.e., how speech is organized into prosodic units (mora, syllable, stress foot, accentual phrase, etc.) in a particular language may affect the extent and directionality of coarticulatory effects. For example, carryover effects may be greater, or smaller than, or equal to anticipatory effects, depending on the way in which speech is organized.

The effect of accent seems to be language specific or dependent on accent type. Fowler (1981) and Magen (1984) showed that unstressed (full) vowels are more susceptible to contextual effects in $F_2$. In Part I, it was suggested that the reduced vowel /ə/ of English may be completely transparent to contextual effects in $F_2$ and thus is extremely variable. This implies that there is a relationship in the magnitude of variability that is inversely related to the degree of stress placed on the segment. That is, vowels are more susceptible to contextual effects in the following order: stressed full vowels < unstressed full vowels < reduced (zero stress) vowels. Thus, stress accent seems to determine the extent of contextual assimilation in English.

Bergem (1993) also observed a similar relationship between the degree and nature of accent and contextual assimilation. He found significant effects of sentence accent and word stress on the extent of vowel reduction in Dutch. For example, while a stressed syllable with a sentence accent was more peripheral in the acoustic vowel space, an unstressed syllable without a sentence accent was more central in formant values. Bergem describes the relationship between vowel reduction and centralization as follows: 'spectral vowel reduction is not a tendency of vowels to centralize, but rather the result of an increased contextual assimilation. The observed centralization of formant patterns is a frequently occurring and probably natural consequence of this assimilation process.'[1]

It was stated on page 153 in Chapter 6 that the extent of centralization observed in different languages as a function of accent may suggest the effect of accent on the extent of contextual assimilation. In Experiment 3, no significant effect of accent was observed on the $F_2$ values of the Japanese vowels /i, e, a, o, u/ though it was suggested that accent may raise $F_1$ values. From this, the accent seems to have little or no effect on the extent of contextual assimilation in Japanese. The experiment in this chapter is explicitly designed to test this.

On the other hand, the different accent types observed in English (stress accent) and Japanese (non-stress accent) may determine the extent and the pattern of vowel variability observed in the two languages in a more global way. English

[1]On the other hand, central vowels are not necessarily variable. According to Bates (forthcoming), the stressed central vowel /3:/ as in the word 'bird' is the least variable vowel of all in the British vowel inventory.

manifests the contrast of targeted and targetless $F_2$ values between full and reduced vowels. This contrast seems to play an important role in organizing speech into prosodic units, namely stress feet in English.[2] (See Chapter 4.) This contrast in vowel quality seems to be an important part of the alternating rhythm of strong and weak syllables. On the other hand, Japanese does not seem to have such a contrast. Presumably, all its vowels are targeted. However, large variability has been reported by previous studies. Magen (1984) notes that the effects of the contextual segments extend farther in time across a Japanese vowel when compared to an English (unstressed) full vowel. Thus, Japanese vowels seem to be more transparent than English full vowels. The extent of V-to-V effects may be spread more or less evenly over all the vowels in Japanese to compensate for the lack of a targetless vowel. From this, it is suggested that the extent of variability observed in Japanese vowels may be intermediate in degree between that observed for the English full and reduced vowels. The experiment in this chapter attempts to test this hypothesis.

Such transparency of vowels to contextual effects may be related to the speech production strategy of Japanese. The extent of V-to-V effects across time will be observed in the present experiment. The directionality of V-to-V effects will also be an important issue here. The transparency and opacity of consonants in V-C-V coarticulation will also be studied using the consonants /b/ and palatalized /$b^j$/ as the middle consonant.

## 7.2 Experiment 4a: Effect of accent

- Hypothesis: Accent does not affect the extent of vowel variability as a function of contexts in Japanese.

[2]The way I see stress foot is different from Selkirk's or Abercrombie's definitions (see Footnote 5 in Chapter 5). According to my view, a stress foot is organized from the onset of a stressed vowel to (and including) the onset of the next stressed vowel.

### 7.2.1 Methods

#### Materials

The VsekV sequences were embedded in the sentences below with and without accent on the middle vowel /e/. The vowel /e/ was chosen as a target vowel as it is known to be less affected by the vowel devoicing process. (See 5.5 (p 106) for more discussion on vowel devoicing.) However, there were a few cases where the /e/ was deleted. The contextual vowels were /i, e, a, o, u/. Some of the contextual vowels /i/ and /u/ were devoiced. In particular, all the /u/'s in the second syllable of the word *sekuto-shugi* (sectarianism) were devoiced for all of the speakers. For the 'oCeCo' sequences, /t/ was used instead of /s/ for the first consonant since real words could not be embedded into the sequence 'oseko'.

1. ise'ki: Ano aka**i se'ki** ga shitei-se'ki desu. (Those red seats are reserved seats.)

2. iseki: Tadashi**i seki**jun-ni shitagat-te chakuseki-shi-te-kudasa'i. (Please be seated according to the right seating order.)

3. ese'ke: Koo-shi-t**e se'ke**n no kaze ni ataru-no-mo yo'i-koto-da. (It is good for him to face the trials and tribulations of the world.)

4. eseke: Koo-shi-t**e seke**n-tei-ba'kari ki-ni-suru-no-wa koma'ru. (It is not good to constantly worry about what the other people think of you.)

5. ase'ka: Amazon-gawa g**a se'ka**i-de ichiban naga'i kawa' desu. (The Amazon is the longest river in the world.)

6. aseka: Mishima Yukio w**a seka**i-teki-ni yuumei-na nihon-no sakka desu. (Yukio Mishima is a Japanese writer who is well-known over the world.)

7. ote'ko: Kon**o te'ko**-no gen'ri-wa iroiro-na bu'nya-de riyoo-o sare-te-iru. (The physics of lever is put to use in various fields.)

8. oteko: Wa'ga-sha-de-wa kono bu'mon-n**o teko**-ire-ga hitsuyoo desu. (It is necessary to put more emphasis in this section in our company.)

9. use'ku: Ka're-wa i'tsumo se'kaseka-shi-te-iru-ga, anna-huu-ni ki-ga su'g**u** **se'ku** no-wa karada-ni yo'ku-nai. (He is always restless, but it is not good for health to be so impatient the way he is.)

10. useku: Kono-yoo-ni su'g**u** **seku**to-shugi-ni hashi'ru-no-wa yo'ku-nai-koto-da. (It is not good to readily get into sectarianism.)

### Speakers

Seven male Japanese speakers participated in this experiment. Of them, six were recruited as paid volunteers at International Christian University in Tokyo. They were undergraduate students at the time of the recording. HI, JK, MT, TK and TM are originally from Tokyo. JN is originally from Nagano, a prefecture to the North of Tokyo. Some of them had lived abroad for some time. HI spent eight months in Canada when he was a high school student. JK spent the age of 16 to 17 in Canada. MT spent the total of six years in the U.K. and Canada from the age of seven to twelve. TK spent seven years in Brazil from the age of four to eleven. JN and TM had never been outside Japan. MI was a visiting fellow at the Department of Cognitive Science, University of Edinburgh, and is slightly older than the other participants. He is originally from Hyoogo, a prefecture in the Western part of Japan, but grew up in Tokyo. He had been living in Edinburgh for a period of one year at the time of the recording. They all speak so called 'Standard Japanese'. MT had some misplacement of accent, but this did not affect the test sequences and his recording was used. There was some technical trouble at the recording for JK and only the first eight repetitions were used for analysis from his data.

### Recordings

For the six subjects recruited in Tokyo, the recording was done at the sound treated recording studio of International Christian University.[3] MI's recording was done in the recording studio at the Department of Linguistics, University of

[3]Thanks are due to Prof. Yoshioka and the recording staff of International Christian University for letting me use the facilities.

| | p-value | | | | | |
|---|---|---|---|---|---|---|
| | F1 | | | F2 | | |
| | Accent | Vowel | $A \times V$ | Accent | Vowel | $A \times V$ |
| HI | 0.0081 | 0.0001 | 0.0053 | 0.0001 | 0.0000 | 0.0012 |
| JK | 0.4902 | 0.0002 | 0.0354 | 0.1430 | 0.0000 | 0.2236 |
| JN | 0.0000 | 0.0008 | 0.0022 | 0.0004 | 0.0000 | 0.0004 |
| MT | 0.0000 | 0.0007 | 0.0000 | 0.6629 | 0.0000 | 0.4801 |
| TK | 0.0000 | 0.0000 | 0.0000 | 0.1086 | 0.0000 | 0.4097 |
| TM | 0.0014 | 0.0000 | 0.0000 | 0.0205 | 0.0000 | 0.2281 |
| MI | 0.0000 | 0.0000 | 0.0000 | 0.4860 | 0.0000 | 0.0136 |

**Table 7.1.** The level of significance of the results of ANOVAs. The table shows the effects of the accent and vocalic context and their interaction on the first and second formant values of the vowel /e/ for each speaker.

Edinburgh.

#### Analyses

The sentences for analysis were sampled at 16 kHZ into a UNIX SUN workstation with WAVES speech analysis facilities. Formant values were obtained by running the FORMANT program for LPC analysis with 25 ms cos**4 window moving in 5 ms steps. Statistical analyses were performed with the vowel midpoint frequency values. See 3.2.1: p 42 for a more detailed description of the analysis method.

### 7.2.2 Results and discussion

The mean $F_1/F_2$ values of the vowel /e/ across the two accent conditions in the Vs_kV context were 389/2002 Hz (HI), 383/1888 Hz (JK), 362/1975 Hz (JN), 350/2055 Hz (MT), 315/2135 Hz (TK), 425/1937 Hz (TM) and 285/1915 Hz (MI). The $F_1$ values observed in this experiment are generally lower than the results of the other studies (Cf. Table 5.1).

Two-way ANOVAs were performed with the grouping factors of accent and vocalic context on the $F_1$ and $F_2$ values of the vowel /e/ for each speaker. Table 7.1 shows the results. The main effect of accent was significant on $F_1$ except for JK. For $F_2$, the main effect of accent was significant for three speakers only, HI, JN and TM. The main effect of vocalic context was highly significant for both $F_1$

| | F1 | | F2 | |
|---|---|---|---|---|
| | Accented | Unaccented | Accented | Unaccented |
| HI | 396.94 | 380.63 | 1952.1 | 2048.2 |
| JK | 380.42 | 384.95 | 1887.6 | 1835.8 |
| JN | 376.32 | 345.82 | 1952.5 | 1999.4 |
| MT | 367.88 | 332.06 | 2051.4 | 2058.9 |
| TK | 342.60 | 287.70 | 2113.7 | 2156.6 |
| TM | 431.86 | 417.74 | 1914.9 | 1958.3 |
| MI | 309.68 | 260.23 | 1910.9 | 1919.6 |

**Table 7.2.** The mean $F_1$ and $F_2$ values of the vowel /e/ as a function of accent for each speaker.

and $F_2$ for all of the speakers.

Table 7.2 shows the mean $F_1$ and $F_2$ values as a function of accent for each speaker. The mean $F_1$ values are higher for accented tokens than for unaccented ones for the six speakers who showed significant difference. This observation supports the results of Experiment 3. There may be a generally more open articulation for accented tokens. For the three speakers who showed significant difference in $F_2$ value as a function of accent, the $F_2$ values were higher for unaccented tokens. The differences are 96.1 Hz, 46.9 Hz and 43.4 Hz for HI, JN and TM.

Three-way ANOVAs were also performed with the pooled data across speakers. The independent variables were accent, vocalic context and speaker. All the main effects and interactions were significant for $F_1$. The effect of accent was significant by $F(1{,}592) = 126.58$, $p < 0.0001$. The result of the post hoc scheffe test for the accent $\times$ speaker interaction showed significant effects of accent on $F_1$ for only MI and TK. For $F_2$, all the main effects and the interactions were significant except for the accent $\times$ vocalic context interaction. The effect of accent on $F_2$ was significant by $F(1{,}592) = 10.94$, $p = 0.0009$. The result of the post hoc scheffe test for the accent $\times$ speaker interaction showed no significant effect of accent for any of the speakers on his own. As the scheffe test is more conservative, these results may be more reliable and fair. It is also interesting to note that there was no significant interaction between accent and vocalic context.

That is, when the vowels in the same vocalic context were compared as a function of accent, unaccented vowels invariably had higher $F_2$ values than accented vowels. The results of the post hoc scheffe test showed that the difference as a function of accent was significant for the vowels in the contexts of /a/ and /o/. These results seem to suggest that unaccented vowels have higher $F_2$ values than accented vowels though the difference is small and statistically only marginally significant.

Figure 7.1 and 7.2 illustrate the spread of the vowel /e/ as a function of vocalic contexts across the two accent conditions. For $F_1$, the effects are systematic for HI, JK and TM. The mean $F_1$ values are lowest in the context of /i/ and highest in the context of /a/ while the mean $F_1$ values for the vowels /e, o, u/ come in between. For the other speakers the patterns vary from one speaker to another. The $F_2$ values of the vowel /e/ vary systematically as a function of vocalic contexts for every speaker. In the contexts of the front vowels /i/ and /e/, the $F_2$ values are higher, while in the contexts of the back vowels /a, o, u/, the $F_2$ values are lower. The differences in the mean $F_2$ values of the vowel /e/ as a function of the contextual vowels /i/ and /u/ are 232 Hz (HI), 363 Hz (JK), 186 Hz (JN), 223 Hz (MT), 277 Hz (TK), 245 Hz (TM) and 206 Hz (MI) at the vowel midpoint for the Vs_kV context. These differences are relatively large. This may be due to the effects of the following /k/. In Experiment 1 on the vowel variation of /ə/, it was observed that there is a greater spread in $F_2$ as a function of vocalic contexts in the consonantal context of /k/.[4]

Figure 7.3 shows the mean $F_1$ and $F_2$ trajectories as a function of vocalic contexts across seven speakers. The $F_1$ trajectories are declining at the vowel offset. The trajectories of JK and MT showed a decline from the onset to the offset. The other speakers showed a rise at the vowel midpoint and then a decline towards the offset. The $F_2$ trajectories of the seven speakers looked more or less the same. The trajectories in the front vowel context and those in the back vowel context are clearly separated. The $F_2$ trajectories diverge towards the offset.

[4] In Experiment 1, the differences in the mean $F_2$ values of the English /ə/ as a function of the contextual vowels /I/ and /u/ were 132 Hz, 219 Hz and -6 Hz for AH, MB and DG in the Vp_pV context. In the Vt_tV context, the differences were -52 Hz, 54 Hz and 27 Hz for AH, MB and DG respectively. In the Vk_kV context the differences were 470 Hz, 227 Hz and 453 Hz for AH, MB and DG. These differences were observed at the vowel midpoint.

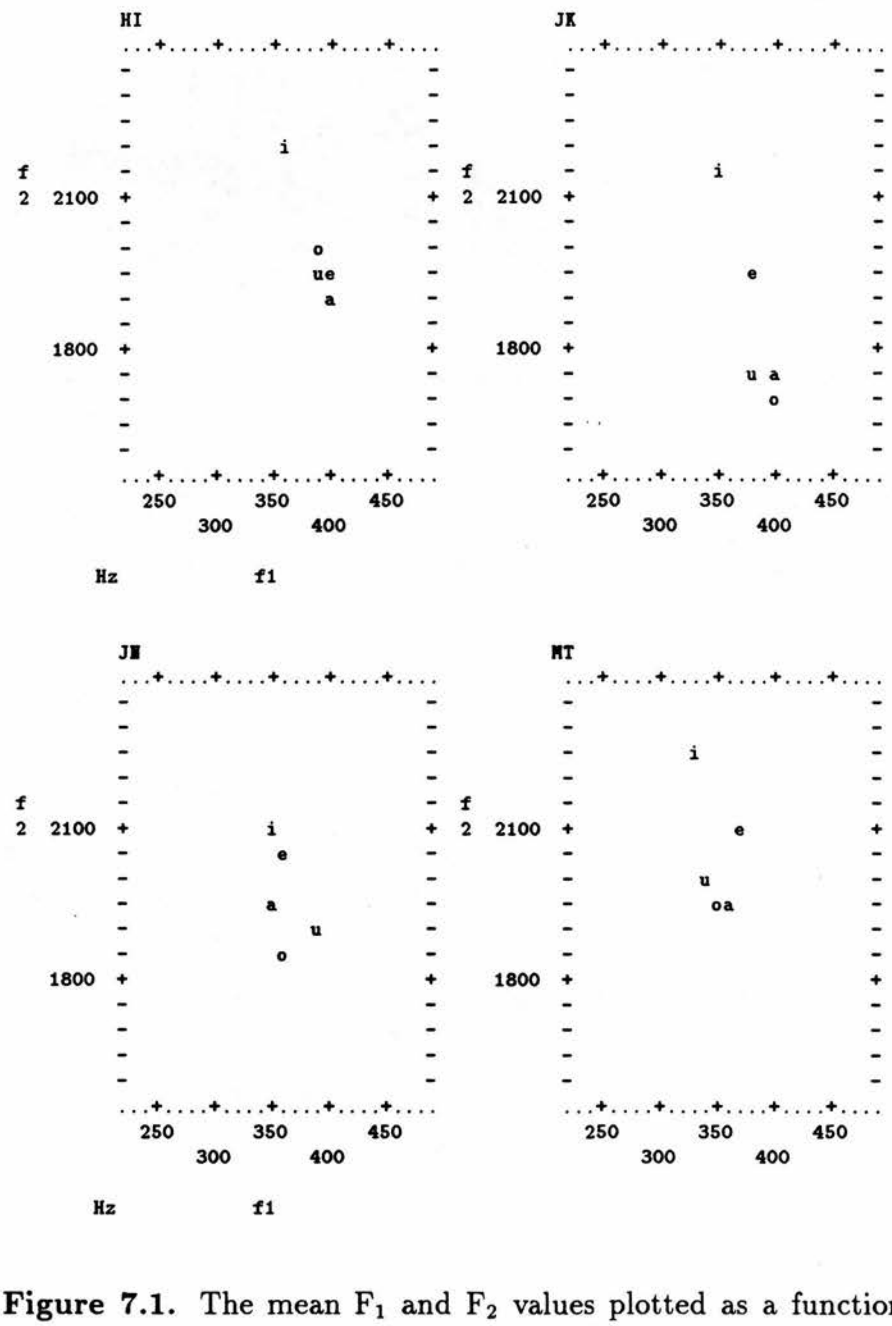



**Figure 7.1.** The mean $F_1$ and $F_2$ values plotted as a function of the vocalic contexts for the speakers HI, JK, JN, and MT.

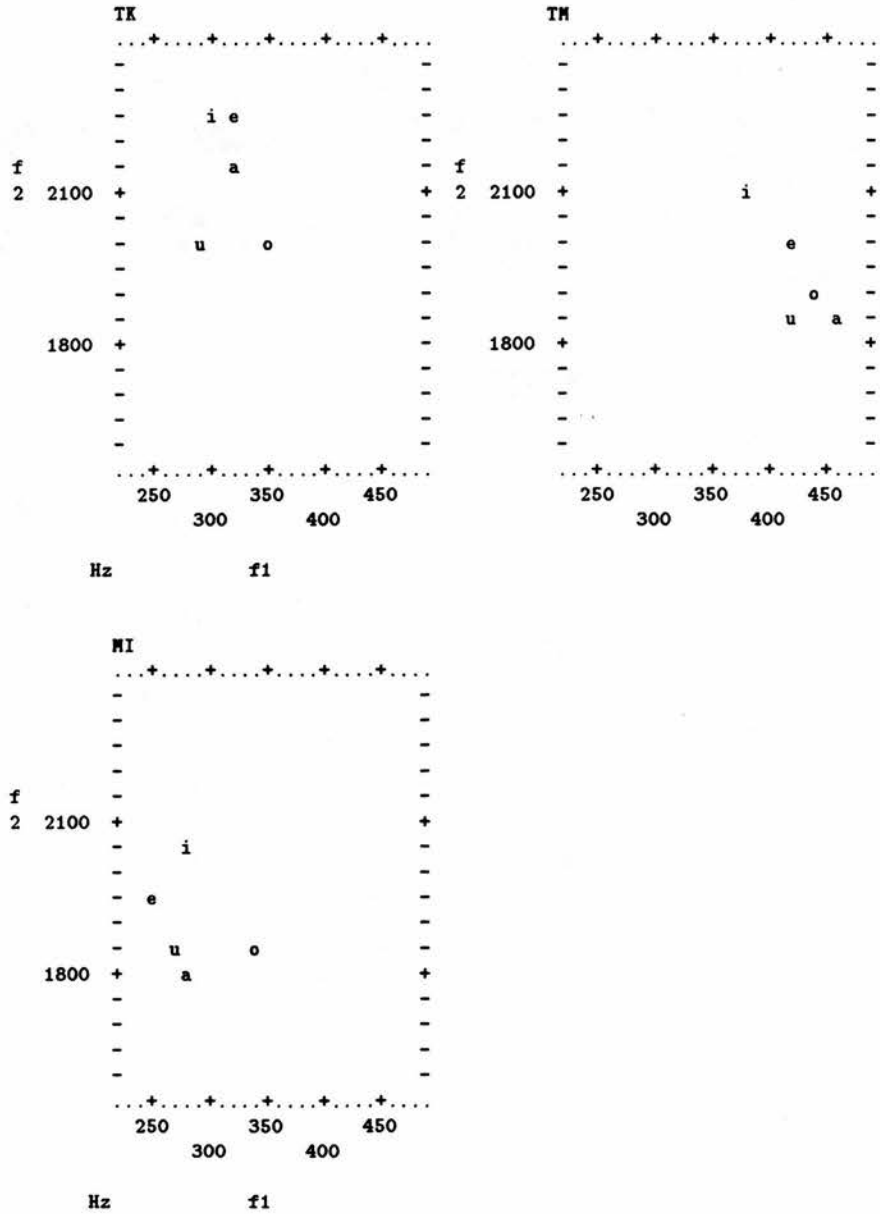



**Figure 7.2.** The mean $F_1$ and $F_2$ values plotted as a function of the vocalic contexts for the speakers TK, TM and MI.

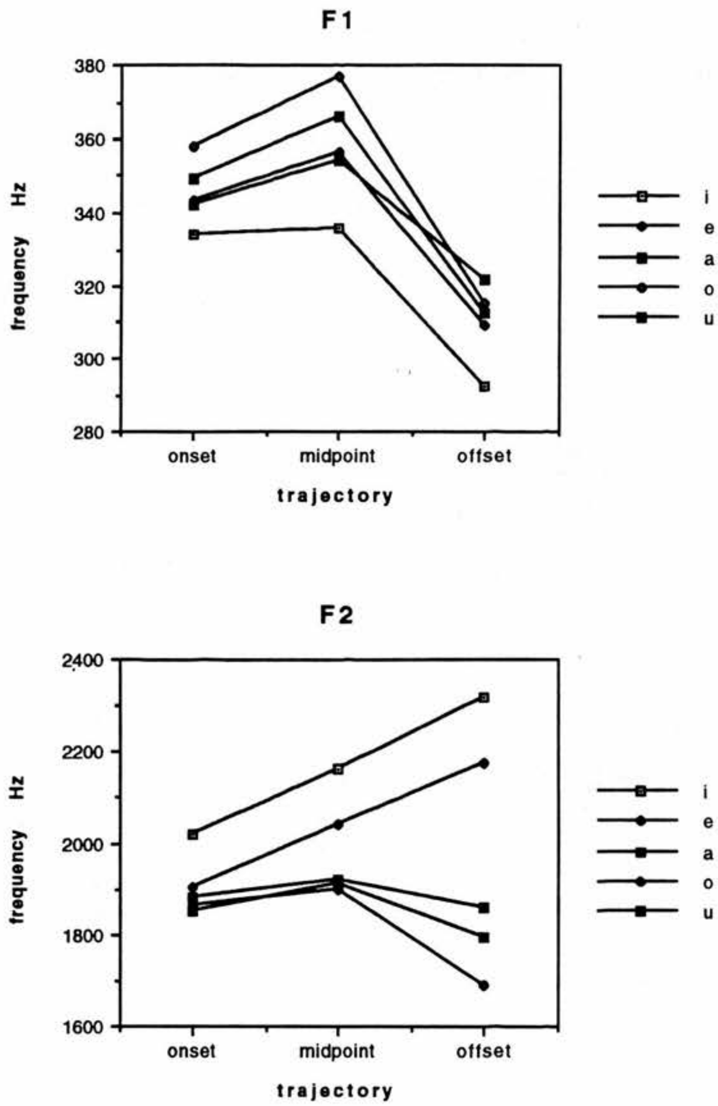



**Figure 7.3.** The mean $F_1$ and $F_2$ trajectories of the vowel /e/ as a function of vocalic contexts across seven speakers.

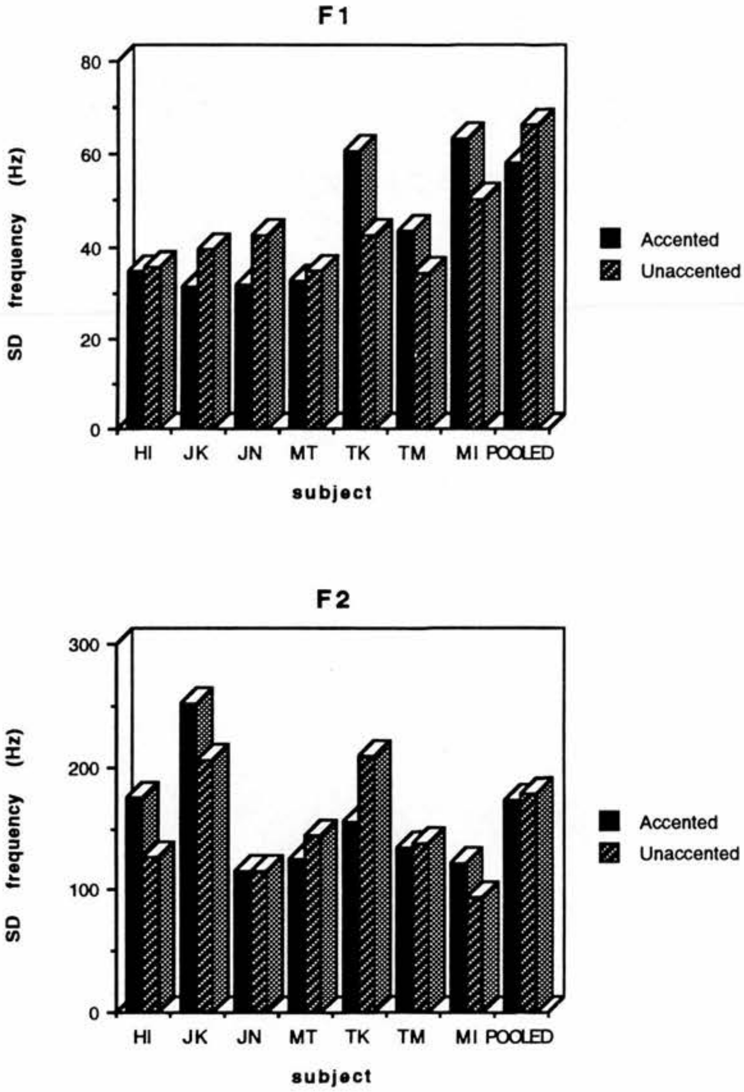



**Figure 7.4.** The standard deviations for the accented and unaccented tokens of the vowel /e/ for $F_1$ and $F_2$ for each speaker and the pooled data.

Thus, the spread in trajectories as a function of vocalic contexts, front vs. back, grows more pronounced from the onset, to the midpoint to the offset. This seems to suggest that anticipatory V-to-V effects are stronger in Japanese than carryover effects. This is in accordance with the observation made by Magen (1984).[5] The differences in the mean $F_2$ values of the vowel /e/ as a function of contextual vowels /i/ and /u/ at the vowel offset are 325 Hz (HI), 586 Hz (JK), 315 Hz (JN), 531 Hz (MT), 605 Hz (TK), 470 Hz (TM) and 427 Hz (MI). These differences are quite striking. This great anticipatory effect may, in part, be conditioned by the following consonant /k/. This is because /k/ itself is assimilated in the place of articulation with the following vowel and thus is transparent to the vocalic effects in [Backness]. (See also Footnote 4.) On the other hand in Experiment 5 below, the anticipatory V-to-V effects were observed to be stronger than carryover effects in the Vp_pV and Vt_tV contexts as well, though the effects were more pronounced in the Vk_kV context. Thus, it seems that in general strong anticipatory effects are observed in V-to-V coarticulation in Japanese.[6]

Figure 7.4 shows the standard deviations (SD) for the accented and unaccented tokens in $F_1$ and $F_2$ for each speaker and the pooled data. The standard deviations for the pooled data were 58 Hz for accented and 66 Hz for unaccented tokens in $F_1$. For $F_2$, the standard deviations were 173 Hz for accented and 178 Hz for unaccented tokens. The differences in SD are quite small between accented and unaccented tokens for both $F_1$ and $F_2$. For $F_1$, TK, TM and MI showed higher SDs for the accented tokens, while the other speakers had higher SDs for the unaccented tokens. For $F_2$, HI, JK and MI had higher SDs for the accented tokens, while the other speakers had higher SDs for the unaccented tokens. This

[5]In the present experiment, the following vowel and the target vowel always belonged to the same word, while there was a word boundary between the preceding vowel and the target vowel /e/ as in the sequence *akai se'ki*, and sometimes a pause was observed at the boundary. In Magen's study, however, the VbV nonsense words were used, and both the preceding and following vowel belonged to the same word as the target vowel, and yet anticipatory coarticulation was stronger.

[6]For schwa, the differences in the mean $F_2$ values as a function of the contextual vowels /I/ and /u/ increased from the onset to the offset in the Vk_kV context: 427 Hz $<$ 470 HZ $<$ 543 Hz (AH), 270 Hz $>$ 227 Hz $<$ 430 Hz (MB), and 412 Hz $<$ 453 Hz $<$ 528 Hz (DG). In the Vp_pV and Vt_tV contexts, patterns were more diverse. In the labial context, AH and MB had stronger carryover than anticipatory V-to-V effects: 156 Hz $>$ 132 Hz $>$ 40 Hz (AH) and 246 Hz $>$ 219 Hz $>$ 197 Hz (MB) in the Vp_pV context.

seems to suggest that for both $F_1$ and $F_2$, there is no systematic difference in vowel variability as a function of accent. That is, for some speakers accented vowels vary more than unaccented vowels, while the opposite is true for other speakers. Further, no significant effect of accent was observed on the $F_2$ values for JK, MT, TK and MI. That is, there was no difference in the distribution of accented and unaccented tokens for these speakers. From these, we may conclude that the accent plays a null role in affecting the extent of V-to-V coarticulation in Japanese.

### 7.2.3 Summary

1. The main effect of accent was significant on $F_1$ for most speakers. The $F_1$ values were higher for the accented vowels, suggesting a more open articulation.

2. The main effect of accent was significant on $F_2$ for only three out of seven speakers. For those who showed significant effects of accent, the $F_2$ values were higher for the unaccented vowels.

3. Comparing the standard deviations of the $F_1/F_2$ values of the accented and unaccented /e/'s, no systematic effect of accent on vowel variability was observed in either $F_1$ or $F_2$. Accented vowels were more variable for some speakers, while unaccented ones were more variable for others.

4. The effects of the vocalic context were highly significant and systematic for both $F_1$ and $F_2$.

5. Anticipatory V-to-V effects were observed to be stronger than carryover effects in $F_2$ for every speaker.

### 7.2.4 Conclusion

Accent does not seem to affect the extent of V-to-V effects in Japanese. Accent in stress accent languages seems to have an important function in organizing speech into prosodic units. The rhythmic unit of English which spans from stressed vowel to stressed vowel seems to impose a certain pattern of coarticulation on its vowels: the alternation of targeted and targetless vowels in $F_2$. Unlike English, accent in Japanese seems to have only a minor role in organizing speech into prosodic units and therefore is not correlated with the coarticulatory pattern of Japanese.[7] In Japanese accent seems to mark out certain syllables in utterances, introducing a marked pitch configuration by accentual fall and catathesis (see page 102). However, the effect of accent seems to be confined to the domain of pitch apart from a slightly open articulation characterized by higher $F_1$ values.

## 7.3 Experiment 4b: V-to-V effects across time

This experiment serves as a preliminary study to explore the nature of V-to-V coarticulation in Japanese. The results of this experiment will be used as reference for comparison in a later experiment (Experiment 7) where the nature of V-to-V coarticulation in interlanguage will be studied. Three main questions are asked in this experiment.

1. What is the magnitude of V-to-V coarticulation in Japanese? Is the degree of vowel variability intermediate in value between that observed for the English full and reduced vowels?

2. How far does V-to-V coarticulation extend in time in Japanese?

3. In which direction is V-to-V coarticulation in Japanese stronger, carryover or anticipatory?

Having no contrast of targeted and targetless vowels, the degree of vowel variability in Japanese may be intermediate in degree between that observed for the English schwa and the English full vowels. While schwa is completely targetless, the English full vowels in general may be more targeted or hyperarticulated than vowels in other languages where such a contrast is lacking. This hypothesis will be tested in this experiment.

The results of Magen's study (1984) suggest that the effects of the transconsonantal vowel may extend far into a vowel in Japanese. In the present study,

[7] See page 103 for the description of accent in Japanese.

the strength of V-to-V effects across time will be studied by observing the effects of the transconsonantal vowel at the onset, midpoint and offset of the vowel and beyond.

Magen also observed stronger anticipatory effects than carryover effects in Japanese. The results of Experiment 4a above also support this. The asymmetry in the strength of L-to-R and R-to-L effects in V-to-V coarticulation will be studied using asymmetric as well as symmetric contexts of Vb_bV.

## 7.3.1 Methods

### Materials

The VbabV sequences were embedded in natural sentences in Japanese with the contextual vowels of /i/ and /a/. Both symmetric and asymmetric contexts were considered resulting in the four different sentences below. These sentences were repeated five times in a randomized order by each speaker. For the reference English data, see page 76.

1. ibabi: Mukashi **Babi**ronia-to-iu kuni-ga arima'shita. (Once there was a nation named Babylonia.)
2. ibaba: So'to-dewa o'tsh**iba-ba**'kari kasakasa-to oto'-o ta'te-te-imasu. (Outside, leaves are falling making rustling sounds.)
3. ababi: Se'isho-niw**a** **Babi**ronia-no-koto'-ga iroiro ka'ite-arimasu. (In the bible, there are stories about Babylonia.)
4. ababa: Kawa'-niwa k**a'ba-ba**'kari-de-na'ku kiken-na wa'ni-mo imasu. (In the river there are dangerous crocodiles as well as hippos.)

### Speakers

Five speakers participated in this experiment. They are KN, MN, SO, KO and SK.[8] All the five speakers spoke Standard Japanese in reading the materials. All

[8]They have also read the sentences with the VbəbV sequences (see 4.3.1) in English. These data will be compared with the present data in a later experiment (Experiment 7) to study the V-to-V coarticulation in interlanguage.

| | a | | |
|---|---|---|---|
| | onset | midpoint | offset |
| pre_v (p) | 0.0000 | 0.0000 | 0.0022 |
| fol_v (f) | 0.0000 | 0.0000 | 0.0000 |
| speaker (s) | 0.0000 | 0.0000 | 0.0000 |
| p × f | 0.0000 | 0.0001 | 0.4173 |
| p × s | 0.3137 | 0.2592 | 0.5679 |
| f × s | 0.0008 | 0.0000 | 0.0000 |
| p × f × s | 0.1361 | 0.2236 | 0.2068 |

**Table 7.3.** The level of significance of the results of ANOVAs. The table shows the main effects of the preceding vowel, following vowel and speaker, and their interaction on the second formant values of the Japanese vowel /a/ across five speakers.

of them were living in Edinburgh at the time of the recording. The data for the English /ə/ and /æ/ are a subset of data used in Experiment 2. As the numbers of subjects for the English and Japanese data are unequal, the middle three tokens from the five repetitions were taken from each English speaker's data to make the number of observations between the English and Japanese data more or less equal.

**Recording and analyses**

The recording was done in a sound treated studio at the Department of Linguistics, University of Edinburgh. For the analysis method, see 3.2.1: p 42. Statistical analyses were performed with the pooled data across speakers.

### 7.3.2 Results and discussion

Three-way ANOVAs were performed with the $F_2$ values of the Japanese vowel /a/ at the onset, midpoint and offset of the segment as dependent variables. The independent variables were the preceding vowel, following vowel and speaker. Statistical analyses were performed with the pooled data across speakers. The main effects of the preceding vowel, following vowel and speaker were significant at all the three points of the segment (Table 7.3). Significant interactions were observed between the preceding and the following vowel at the onset and midpoint of the vowel. Significant interactions were also observed between the following vowel and speaker across the segment. Speakers showed different extent of variability as a function of the following vowel (Appendix A).

Figures 7.5 to 7.7 show the mean $F_2$ trajectories of the Japanese vowel /a/ in different Vb_bV contexts.[9] Tables 7.4 to 7.8 show the differences in $F_2$ values as a function of the contextual vowels /ɪ/ (/i/) and /æ/ (/a/ or /ə/) for the English /ə/, the Japanese /a/ and the English full vowel /æ/ in the comparable Vb_bV context. The asterisks indicate that the difference has reached the significant level of $p < 0.05$.[10]

Figure 7.5 shows the $F_2$ trajectories for the English /ə/, the Japanese /a/ and the English /æ/ in the symmetric context. Both /ə/ and the Japanese /a/ show great effects of the contextual vowels compared to the English full vowel /æ/. Table 7.4 shows that for the English /ə/, carryover effects are stronger. The differences as a function of vocalic contexts decrease towards the offset for /ə/. For the English /æ/ and the Japanese /a/, the difference is slightly greater at the offset than at the onset. The differences in frequency Hz as a function of contextual vowels at the vowel onset and midpoint are indeed intermediate in value for the Japanese vowel /a/ compared to the English /ə/ and the full vowel /æ/ as predicted by the hypothesis above.

However, the situation seems to be slightly more complicated when the asymmetric contexts are considered. First of all, let us look at the difference between carryover and anticipatory effects in Japanese. Figures 7.6(a) and 7.7(a) are

[9] Each data point in these figures is the mean of 5 repetitions × 5 speakers. In some cases, there were less than 25 observations per cell due to spurious formant tracking. The smallest number of observations in a single cell mean is 21. The reference figures for the English /ə/ and /æ/ are taken from Figure 4.2.

[10] Post hoc scheffe tests were performed for the interaction between the 2 contextual vowels (/I/ or /i/ vs. /æ/, /ə/ or /a/), the 3 points of measurement (onset, midpoint and offset) and the 3 affected vowels (the English /ə/ and /æ/ and the Japanese /a/). There are 18 cell means to compare. There are unequal numbers of subjects for the English and Japanese data (8 vs. 5). In order to make the number of observations per mean more or less equal, the middle three tokens from the five repetitions were taken from each English speaker's data making the total observations per mean 3 repetitions × 8 speakers = 24 observations. For the Japanese tokens, there are 5 speakers × 5 repetitions = 25 observations per mean. The harmonic mean of the cell sizes is used for the scheffe test.

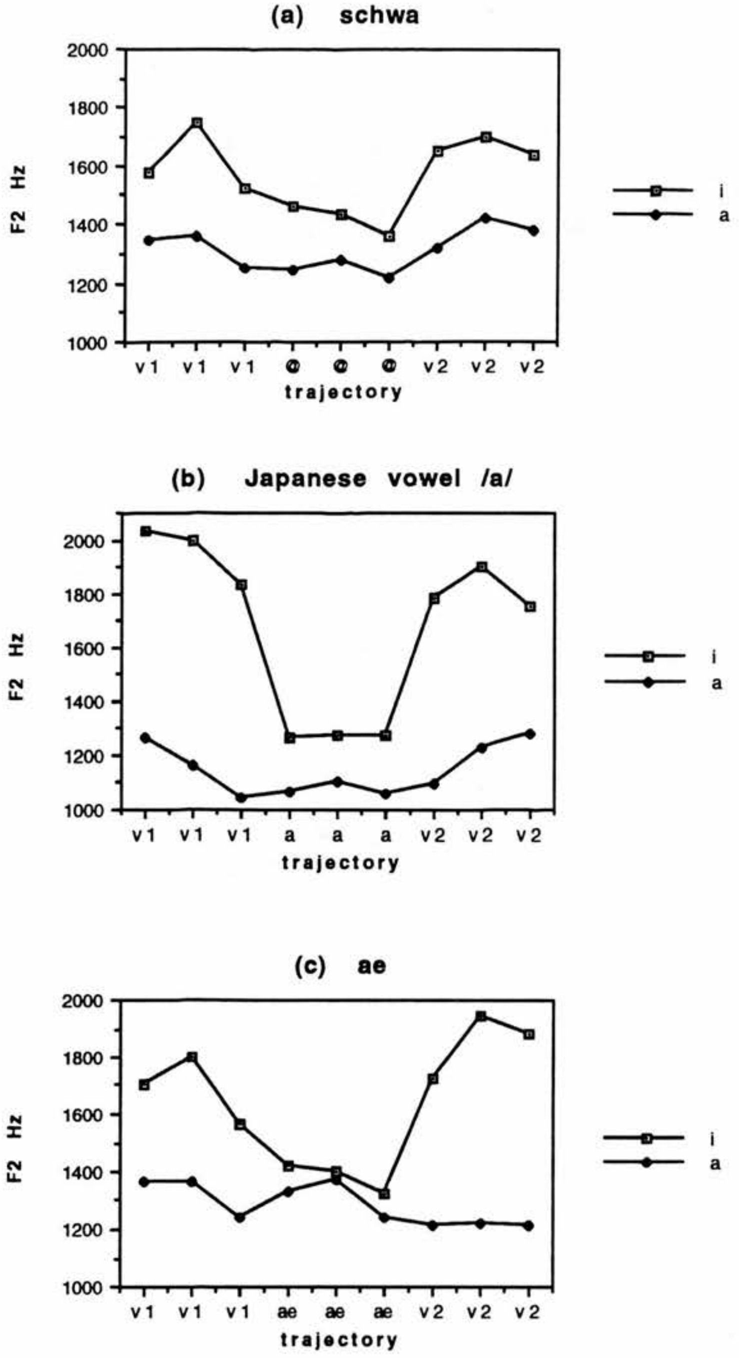
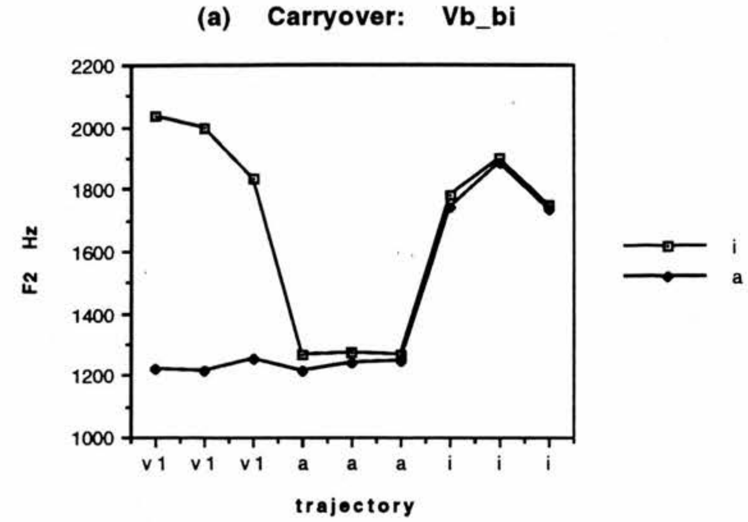



**Figure 7.5.** The mean $F_2$ trajectories for the English /ə/, the Japanese vowel /a/ and the English full vowel /æ/ in the symmetric Vb_bV context. The vocalic contexts are either /ɪ/ and /æ/ (/ə/) or /i/ and /a/.



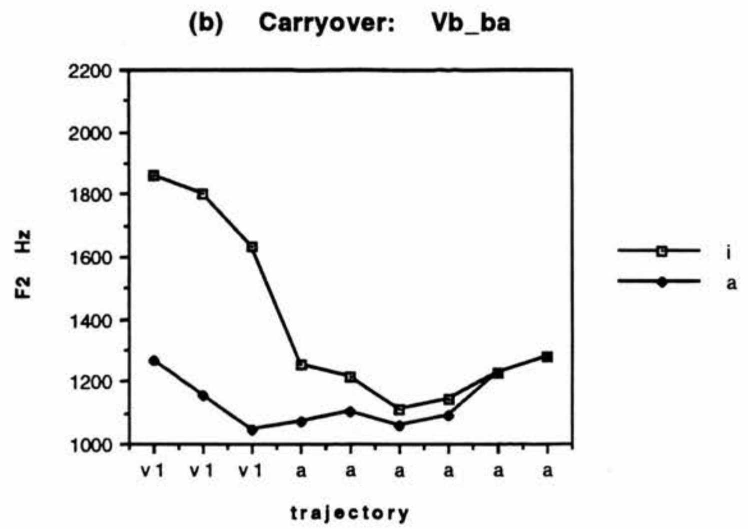




**Figure 7.6.** The mean $F_2$ trajectories for the Japanese vowel /a/ in the asymmetric Vb_bV contexts. The preceding vowels are /i/ and /a/.

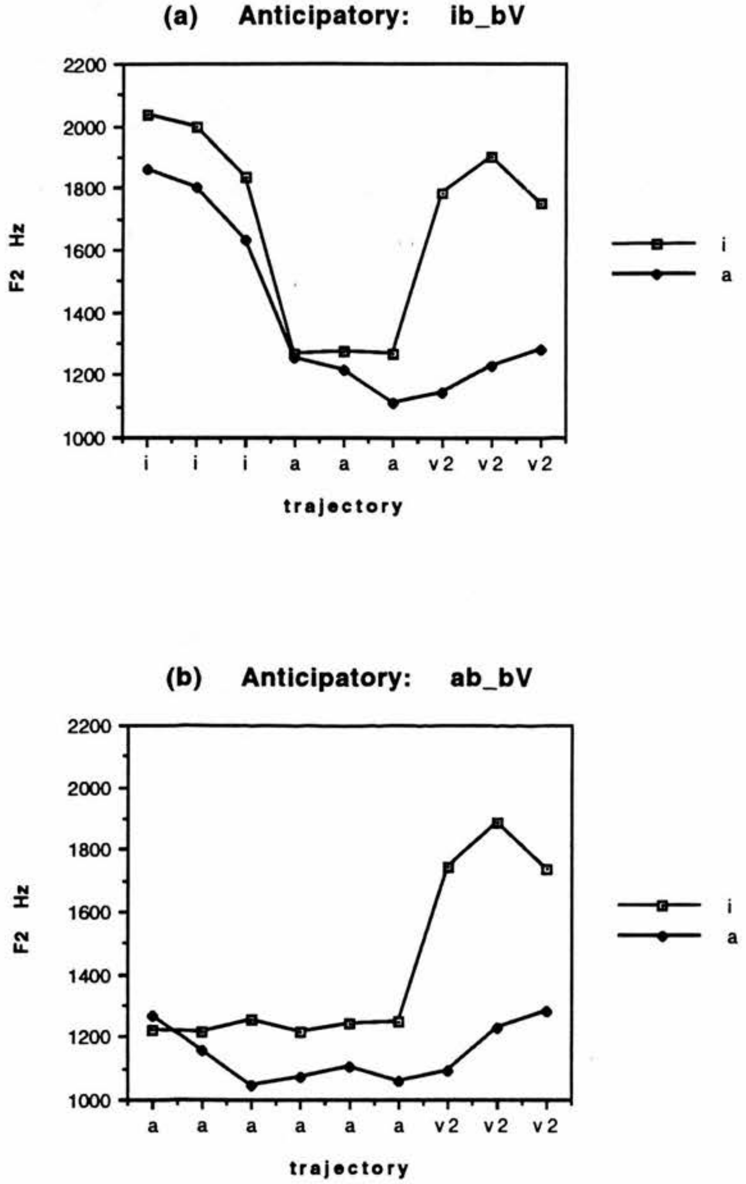



**Figure 7.7.** The mean $F_2$ trajectories for the Japanese vowel /a/ in the asymmetric Vb_bV contexts. The following vowels are /i/ and /a/.

somewhat mirror images of one another. So are Figures 7.6(b) and 7.7(b). When they are compared across pages, it becomes clear that anticipatory effects are stronger. In the Vb_bi and ib_bV contexts, carryover effects were statistically not significant, while anticipatory effects were significant at the vowel offset (Tables 7.5 and 7.7). In the Vb_ba and ab_bV contexts carryover effects were significant at the onset and anticipatory effects were significant at the offset (Tables 7.6 and 7.8). The differences as a function of vocalic contexts were greater for anticipatory than carryover effects.

When Figures 7.6(a) and (b) are compared, stronger carryover effects are observed in (b) where the third vowel is also /a/. Similarly, when Figures 7.7(a) and (b) are compared, stronger anticipatory effects are observed in (b) where the first vowel is /a/. Differences were observed throughout the segment for carryover effects in the Vb_ba context (Table 7.6) and for anticipatory effects in the ab_bV context (Table 7.8) where the segment beyond the middle vowel is also /a/. The effects of $V_1$ were observed at the onset of $V_3$ when $V_3$ was also /a/ (Figure 7.6(b)), and the effects of $V_3$ were observed at the offset and midpoint of $V_1$ when the $V_1$ was /a/ (Figure 7.7(b)). This seems to suggest that when there is a sequence of two identical vowels in VCVCV utterances, the middle vowel becomes more transparent for both carryover and anticipatory effects in that the effects of the transconsonantal vowel penetrate right through it into the other transconsonantal vowel.[11] The mean duration from the offset of $V_1$ to the onset of $V_3$ in the present data is 274.586 ms. The above results suggest that the effect of the vowel extended as far as 275 ms for the carryover V-to-V effects and even farther for the anticipatory effects under certain conditions in Japanese.

For the carryover effects in the Vb_bi (Vb_bɪ) context, the English /ə/ showed greater effects than the Japanese vowel /a/. The English /æ/ showed similar extent of carryover effects as the Japanese /a/. For the carryover effects in the Vb_ba context, the Japanese vowel /a/ showed almost the same magnitude of carryover effects as does the English schwa because of the effect of the /a/ in $V_3$. For the anticipatory effects in the ib_bV context, the Japanese vowel /a/ showed great effect at the offset though such effects did not extend across the segment.

[11]It is possible that this feature is particularly strong for the vowel /a/. Further tests are needed to see what would happen with the other four vowels of Japanese.

| Symmetric: Vb_bV | | | | |
|---|---|---|---|---|
| language | vowel type | onset | midpoint | offset |
| English | ə | *215.3 | *180.4 | 161.0 |
| Japanese | a | *196.0 | *167.3 | *214.8 |
| English | æ | 73.9 | 23.8 | 91.2 |

**Table 7.4.** The difference in $F_2$ value (Hz) as a function of the contextual vowels /ɪ/ (/i/) and /æ/ (/a/ or /ə/) for the English /ə/, the Japanese vowel /a/ and the English full vowel /æ/ in the symmetric Vb_bV context. The symbol * indicates that the difference has reached a statistically significant level ($p < 0.05$).

| Carryover: Vb_bi | | | | |
|---|---|---|---|---|
| language | vowel type | onset | midpoint | offset |
| English | ə | *188.9 | 106.6 | 96.7 |
| Japanese | a | 52.2 | 31.1 | 20.6 |
| English | æ | 68.5 | 30.1 | 9.8 |

**Table 7.5.** The difference in $F_2$ value (Hz) as a function of the preceding vowels /ɪ/ (/i/) and /æ/ (/a/) for the English /ə/, the Japanese vowel /a/ and the English full vowel /æ/ in the asymmetric Vb_bi context. The symbol * indicates that the difference has reached a statistically significant level ($p < 0.05$).

| Carryover: Vb_ba | | | | |
|---|---|---|---|---|
| language | vowel type | onset | midpoint | offset |
| English | ə | *159.4 | 103.3 | 57.6 |
| Japanese | a | *187.6 | 111.0 | 52.9 |
| English | æ | 70.3 | 13.7 | -33.5 |

**Table 7.6.** The difference in $F_2$ value (Hz) as a function of the preceding vowels /ɪ/ (/i/) and /æ/ (/a/) for the English /ə/, the Japanese vowel /a/ and the English full vowel /æ/ in the asymmetric Vb_ba context. The symbol * indicates that the difference has reached a statistically significant level ($p < 0.05$).

| Anticipatory: ib_bV | | | | |
|---|---|---|---|---|
| language | vowel type | onset | midpoint | offset |
| English | ə | 55.9 | 77.1 | 103.4 |
| Japanese | a | 8.4 | 56.3 | *161.9 |
| English | æ | 3.6 | 10.1 | 124.7 |

**Table 7.7.** The difference in $F_2$ value (Hz) as a function of the following vowels /ɪ/ (/i/) and /æ/ (/a/ or /ə/) for the English /ə/, the Japanese vowel /a/ and the English full vowel /æ/ in the asymmetric ib_bV context. The symbol * indicates that the difference has reached a statistically significant level ($p < 0.05$).

| Anticipatory: ab_bV | | | | |
|---|---|---|---|---|
| language | vowel type | onset | midpoint | offset |
| English | ə | 26.4 | 73.8 | 64.3 |
| Japanese | a | 143.8 | 136.2 | *194.2 |
| English | æ | 5.4 | -6.3 | 81.4 |

**Table 7.8.** The difference in $F_2$ value (Hz) as a function of the following vowels /ɪ/ (/i/) and /æ/ (/a/ or /ə/) for the English /ə/, the Japanese vowel /a/ and the English full vowel /æ/ in the asymmetric ab_bV context. The symbol * indicates that the difference has reached a statistically significant level ($p <0.05$).

| English | AH | MB | DN | JD | JP | mean |
|---|---|---|---|---|---|---|
| /ə/ | 51 | 57 | 43 | 41 | 37 | 46 |
| /æ/ | 74 | 65 | 61 | 83 | 69 | 70 |
| Japanese | KN | MN | SO | KO | SK | mean |
| /a/ | 62 | 78 | 63 | 80 | 65 | 70 |

**Table 7.9.** The mean durations of the English /ə/, /æ/ and the Japanese vowel /a/ for each speaker and across speakers.

The English /ə/ showed the effects of the third vowel throughout the segment. For the English /æ/, the effects of the following vowel seem to be stopped at the offset. For the anticipatory effects in the ab_bV context, the Japanese vowel /a/ showed the greatest effects of the following vowel among the three vowels.

For the Japanese vowel /a/, V-to-V effects are stronger when the direction of the effects is R-to-L, i.e., anticipatory. The effects are further strengthened when the vowel beyond is also another /a/. Thus, the extent of V-to-V coarticulation in Japanese seems to be determined by the direction of the effects and the quality of the vowel beyond the affected segment. On the other hand, carryover effects seem to be stronger than anticipatory effects (in the /b/ context) for the English vowels.[12]

Table 7.9 shows the mean durations for the English /ə/, the Japanese /a/ and the English full vowel /æ/ for each speaker and across speakers.[13] The mean durations for the Japanese vowel /a/ and the English full vowel /æ/ are similar. While the segment durations observed in the present study were similar between the Japanese vowel /a/ and the English full vowel /æ/, they manifested quite different coarticulatory patterns. The extent of variability as a function of contexts was far greater for the Japanese /a/ than for the English /æ/. Also the direction of the coarticulatory effects seemed to determine the strength of the effects quite differently between the two languages. While the Japanese /a/ favoured anticipatory effects, the English /æ/ showed similar strength between

[12]In the /k/ context, anticipatory effects were observed to be stronger than carryover effects on schwa (see Table 3.6).

[13]For the English data, the first five subjects in Table 4.7 are shown in Table 7.9.

carryover and anticipatory effects. The results of the ANOVAs in Chapter 4 suggest that carryover effects may extend farther into a segment than anticipatory effects for the vowel /æ/ (see Table 4.1). The V-to-V effects extend through the segment in Japanese, but for the English full vowel /æ/, the effects are observed only at the onset and midpoint for carryover and at the offset for anticipatory coarticulation. It seems that the extent and the pattern of coarticulation is largely determined at a higher level in the speech production planning. Though the temporal factor may interact with this plan at the time of the execution, the general pattern may not vary so largely.

In the present study, strong V-to-V effects were observed on the Japanese vowel /a/. However, other vowels in Japanese may not be as contextually variable as /a/. Magen (1984) reports that while the amount of coarticulation in /a/ and /i/ was roughly the same in English, the vowel /a/ showed much more coarticulation than did the vowel /i/ in Japanese. Recasens (1991) observed little variability in the front vowels /i, e, ɛ/ of Catalan while the back vowels /a, ɔ, o, u/ showed large variability in the linguopalatal contact across different consonantal environments (an EPG study). Kiritani *et al.* (1977) also observed that variation in the positions of the tongue tip pellet were larger for the back vowels /a, o, u/ than for the front vowels /i, e/ in Japanese (an X-ray microbeam study). Perturbations due to the neighbouring consonants introduced confusion among back vowels as far as the tongue configuration was concerned. In such cases the distinction was achieved by the lip and jaw control. The same trend was observed by Choi & Keating (1991) for V-to-V effects. Difference in $F_2$ value due to different transconsonantal vowels was more robust when the affected vowel was /a/ than when it was /i/ for three Slavic languages, Russian, Bulgarian and Polish. Thus, inherent variability of each vowel within a language must be studied in comparison with one another. Further investigations are desired in this area.

It should also be remembered that the quality of the contextual vowels is different between English and Japanese. For example, the mean $F_2$ values of the preceding vowels /ɪ/ and /æ/ in English were 1729.6 Hz and 1362.1 Hz respectively, whereas the mean $F_2$ values of the preceding vowels /i/ and /a/ in Japanese were 1940 Hz and 1188.7 Hz. Thus, the influence is greater for the Japanese vowel /a/ than for the English /ə/ or /æ/, and yet the English schwa

| | | $V_1$ | $V_2$ |
|---|---|---|---|
| English /ə/ | /ɪ/ | 1692.6 | 1712.6 |
| | /æ/ | 1353.5 | 1403.0 |
| English /æ/ | /ɪ/ | 1765.1 (stressed) | 1937.8 (unstressed) |
| | /æ/ | 1370.7 | 1151.2 (/ə/) |
| Japanese /a/ | /i/ | 1940.0 | 1891.2 |
| | /a/ | 1188.7 | 1225.4 |

**Table 7.10.** The mean $F_2$ values at the midpoint of the $V_1$ and $V_2$ of the Vb_bV context for the English schwa, full vowel /æ/ and the Japanese vowel /a/.

showed greater variability as a function of the contextual vowels. On the other hand, the difference in variability observed between the English /æ/ and the Japanese /a/ must be assessed with some reservation. Table 7.10 shows the mean $F_2$ values of the preceding and following vowels for the English /ə/, /æ/ and the Japanese /a/. (See also Figure 7.5.) The difference in $F_2$ values observed between the preceding and following vowels may be due to contextual effects.

## 7.4 Experiment 4c: Effect of secondary articulation

In Japanese there is a distinctive contrast of palatalized and non-palatalized consonants before the three back vowels /a, o, u/: e.g., *ha'ku* (foil) vs. *hyaku'* (hundred) (see page 99). The effects of these palatalized consonants in blocking the V-to-V articulation will be studied in this experiment.

### 7.4.1 Methods

The Japanese vowel /a/ was placed in the VCa context with the preceding consonants of /b/ and /$b^j$/ and the preceding transconsonantal vowels of /i, a, o/ to see the effect of palatalization in blocking carryover V-to-V effects. The six different VCa sequences were embedded in the following sentences. Speaker MN read the sentences 8 times in a randomized order. The recording was done in a sound treated room in the Language Laboratory of the University of Edinburgh. The analyses were performed in the same procedure as in 3.2.1: p 42.

1. iba: Ak**a**i **ba**kuchiku-ga machi'-o kazat-te-iru. (Red firecrackers are decorating the streets.)
2. ib$^j$a: Taka'**i** **bya**ku'dan-no sensu-o kat-ta. (I bought an expensive sandal wood fan.)
3. aba: Kore-**ga** **ba**'ku-to-iu doobutsu desu. (This is the animal named 'baku' (tapir).)
4. ab$^j$a: Yo'ru-no sanpo-**ga** **bja**'kuya-no tanoshimi' desu. (It is a pleasure to have a walk on a bright summer evening.)
5. oba: Bi'ru-n**o** **ba**kuha-de shisho'osha-ga de'-ta. (Lots of people were injured and killed in the explosion of the building.)
6. ob$^j$a: An**o** **bya**'kuya-no huukeiga wa utsukushi'i. (That sketch of the midnight sun is beautiful.)

### 7.4.2 Results and discussion

The mean $F_1$ value for the palatalized /$^j$a/ is 408 Hz while the mean $F_1$ value for the vowel /a/ is 689 Hz. The $F_1$ value is lowered by the palatal secondary articulation. The mean $F_2$ value for the vowel /a/ is 1200 Hz while the mean $F_2$ value for the palatalized /$^j$a/ is 1568 Hz. The $F_2$ value is raised. Two-way ANOVAs were performed with the preceding consonants (/b/ vs. /b$^j$/) and the preceding vowels (/i, a, o/) as the grouping factors. There were significant effects of the preceding consonants on both $F_1$ ($F(1,42) = 20.16$, $p = 0.0001$) and $F_2$ ($F(1,42) = 1050.69$, $p < 0.0001$). No significant effects of the transconsonantal vowels were observed on $F_1$ ($F(2,42) = 0.73$, $p = 0.4883$). Significant effects of the preceding vowels were observed on $F_2$ ($F(2,42) = 12.84$, $p = 0.0001$).

A separate sets of ANOVAs for the /ba/ and /b$^j$a/ were performed with the preceding vowel as independent variable. Significant effects of the preceding vowels were observed on both the $F_2$ values of the vowel /a/ at the onset (F(2,21)

| | | onset | | | midpoint | | | offset | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | i | a | o | i | a | o | i | a | o |
| $F_1$ | b | 427 | 572 | 486 | 600 | 746 | 722 | 558 | 648 | 530 |
| | $b^j$ | 339 | 307 | 319 | 518 | 352 | 355 | 389 | 298 | 273 |
| $F_2$ | b | 1274 | 1128 | 1064 | 1261 | 1191 | 1148 | 1169 | 1206 | 1190 |
| | $b^j$ | 1719 | 1723 | 1716 | 1498 | 1616 | 1590 | 1280 | 1335 | 1309 |

**Table 7.11.** The mean $F_1$ and $F_2$ values at the onset, midpoint and offset of the vowel /a/ as a function of the preceding consonants /b/ and /$b^j$/ and the preceding vowel /i, a, o/.

$= 35.60$, $p < 0.0001$) and midpoint ($F(2,21) = 5.47$, $p = 0.0122$) of the segment in the /b/ context. No significant effects of the preceding vowels were observed on the $F_1$ values at the onset ($F(2,21) = 1.35$, $p = 0.2809$). No significant effects were observed for the vowel /$^j$a/ in the /$b^j$/ context for neither $F_1$ nor $F_2$ at the vowel onset: $F(2,21) = 0.5$, $p = 0.6110$ for $F_1$ and $F(2,21) = 0.02$, $p = 0.9785$ for $F_2$.

Table 7.11 shows the mean $F_1$ and $F_2$ values as a function of the preceding consonants /b/ and /$b^j$/ and the preceding vowels /i, a, o/. The systematic effects of the preceding vowels are observed in both $F_1$ and $F_2$ when the preceding consonant is /b/. However, no systematic effects of the vocalic contexts are observed in the context of the palatalized consonant /$b^j$/.

Figure 7.8 contrasts the transparency and opacity of the consonants /b/ and /$b^j$/ to V-to-V effects. The vowel /a/ shows the effect of the preceding vowels following /b/, but /$b^j$/ almost completely blocks the V-to-V effects.

Öhman (1966) and Choi & Keating (1991) suggest that languages with secondary articulation such as palatalization, e.g., Russian, Polish and Bulgarian, show a smaller magnitude of V-to-V coarticulation. Choi & Keating (1991) compared the extent of V-to-V effects (across plain and palatalized consonants) on English, Russian, Bulgarian and Polish. They found that the magnitudes of V-to-V coarticulation grew smaller for English > Polish > Russian > Bulgarian. For all the four languages, carryover effects were greater than anticipatory effects, probably because the first vowel in the VCV sequence was always stressed in their

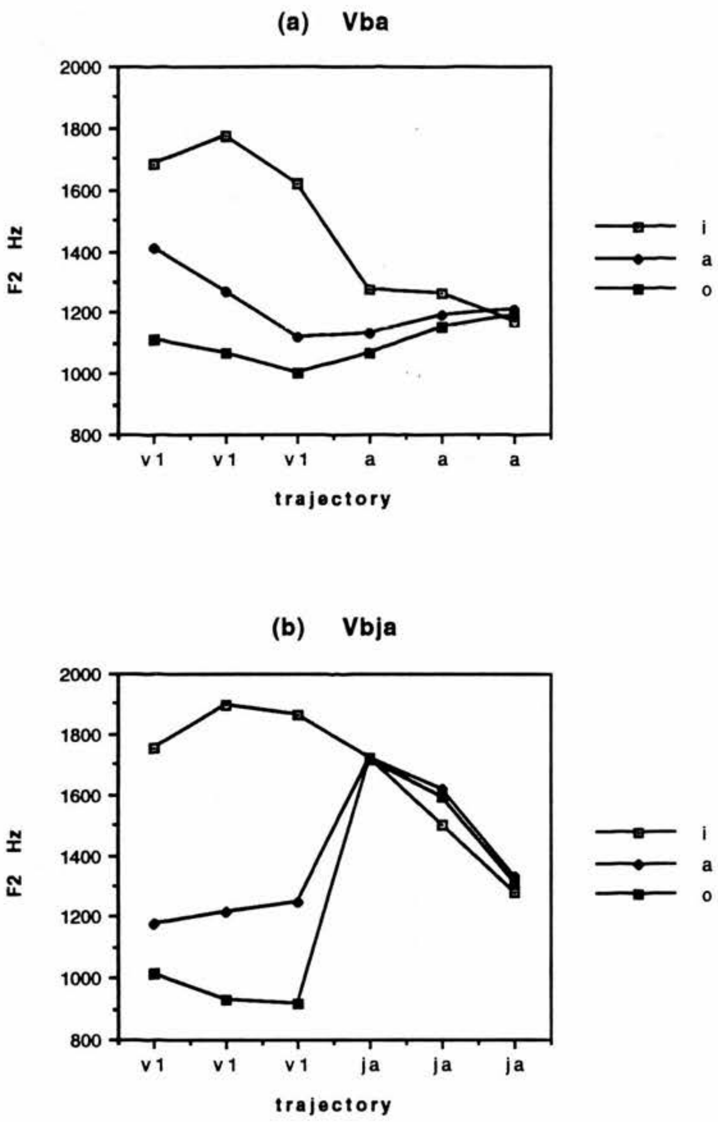



**Figure 7.8.** The mean $F_1$ and $F_2$ trajectories for the vowel /a/ in the contexts of Vba and $Vb^ja$. The preceding vowels are /i, a, o/. Each data point is the mean of 8 observations.

study. Although Japanese has a contrast of palatal vs. non-palatal consonants, it manifests great extent of V-to-V coarticulation as shown in Experiment 4a and b. Secondary articulation on its own may not be sufficient to account for the small magnitude of V-to-V effects observed in the above Slavic languages. The interaction of various unknown constraints with secondary articulation may determine the extent of V-to-V coarticulation in those Slavic languages. For example, Russian is known to have strong vowel reduction. If vowel reduction in Russian is comparable to that observed in English, the reduced vowels in Russian are supposed to be contextually more variable than the unreduced vowels. A more thorough investigation of the overall pattern of coarticulation in Russian is needed to explain the different degrees of coarticulation observed in Russian and Japanese. Another possibility is that the feature [+palatal] is emphasized in these languages by having the non-palatal consonants specified as [+velarized] while in Japanese non-palatal consonants may be unspecified for such features. Keating (1985) argues that consonants in Russian are specified for vowel feature and therefore opaque to V-to-V effects.

## 7.5 Conclusion

The extent of V-to-V effects in Japanese seems to be constrained by a number of conditions, such as the directionality of the effects, the nature of the intervening consonant and the vowel which is beyond the affected segment. These conditions seem to interact with one another to determine the magnitude and the extent in time of the V-to-V effects. In general, strong V-to-V effects were observed for the Japanese vowel /a/. In the symmetric Vb_bV context, the extent of V-to-V effects observed on the Japanese /a/ was intermediate in degree between that of the English /ə/ and the English full vowel /æ/ as predicted by the hypothesis above. However, depending on the conditions discussed above, the Japanese vowel /a/ sometimes showed even greater V-to-V effects than the English /ə/. On the other hand, in the Vb_bi context where the conditions do not favour strong V-to-V effects in Japanese, it showed even smaller extent of V-to-V effects than the English full vowel /æ/. As the number of observations is small in the present study (5 repetitions $\times$ 5 speakers), the results must be assessed with reservation.

However, the extent of V-to-V effects seems to vary systematically as a function of the conditions discussed above. On the other hand, unlike English, accent did not show any systematic effect on the extent of V-to-V coarticulation in Japanese. No systematic difference in the amount of vowel variability was observed as a function of accent. It was suggested that the little correlation observed between accent and contextual assimilation may be due to a minimal role played by the accent in the prosodic organization of Japanese.

# Chapter 8

# C-to-V Coarticulation in Japanese

## 8.1 Introduction

This experiment is designed to study the effects of the immediately adjacent consonants as well as the transconsonantal vowels on the formant frequencies of the vowel /a/ in Japanese. The basic design of the experiment replicates that of Experiment 1 on schwa in order to compare the extent of C-to-V effects between the English schwa and the Japanese vowel /a/.[1] In Experiment 1, it was observed that the effects of the immediately adjacent consonants were far greater than the effects of the transconsonantal vowels on the $F_2$ values of schwa. In the VCəCV contexts, 52.62%, 74.15% and 63.86% of the total variance in the $F_2$ values of schwa could be explained by the consonantal effects for the subjects AH, MB and DG respectively. On the other hand, $V_1$ and $V_2$ together accounted for only 32.65%, 15.98% and 20.73% of the total variance for AH, MB and DG.

This greater effect of the consonantal context seems to be a key factor for the phonetic underspecification of schwa. While the phonetic underspecification of schwa was not clear when only the V-to-V effects were observed, it became clear

[1]The present study is limited in that comparable VC_CV utterances were not observed for the English full vowel /æ/. This is due to the difficulty of placing a full vowel in such a context using natural sentences. Thus, the three-way comparison of the English reduced vowel /ə/, the full vowel /æ/ and the Japanese vowel /a/ could not be made using this particular context.

when the consonantal effects were taken into account. For example, the consonant /t/ blocked V-to-V effects on schwa for the three subjects in Experiment 1. Also, in the context of /p/, the speakers showed different coarticulatory patterns. AH and MB showed carryover V-to-V effects on schwa, but DG showed no V-to-V effects on schwa in the labial context, resulting in his seemingly targeted schwa in Figure 3.5. The labial consonant seems to place a strong constraint on V-to-V coarticulation for DG. Thus, by observing only V-to-V effects, it was not clear whether schwa was targetless or not.

On the other hand, strong effects of V-to-V coarticulation was observed on the Japanese vowel /a/ in the previous chapter. In some cases these effects seemed to be even greater than those observed on schwa. For example, anticipatory V-to-V effects were greater on the Japanese vowel /a/ than on /ə/. Also, in the $V_1CaCV_2$ utterances in Japanese, the quality of $V_2$ (for carryover effects) and the quality of $V_1$ (for anticipatory effects) affected the extent of V-to-V coarticulation. In other words, when either $V_1$ or $V_2$ was also /a/, stronger effects were observed both in magnitude and across time. That is, when $V_1$ was /a/, it already anticipated the effects of $V_2$ across the middle /a/ (significant effects of $V_2$ were observed already at the midpoint of $V_1$), and when $V_2$ was /a/, the effects of $V_1$ was observed right across the middle /a/ to the onset of $V_2$.[2] Thus, the vowel /a/ in Japanese becomes quite transparent to V-to-V effects under certain conditions. This seems to be quite an economical way of speech production as the tongue seems to maintain more or less the same articulatory position throughout the VCV sequence when the two V's are the same.

Strong V-to-V effects suggest less constraint from the intervening consonant. It is possible that consonants place less constraint on V-to-V coarticulation in Japanese resulting in small C-to-V effects.

[2]The results of the t-tests, $p < 0.05$.

## 8.2 Experiment 5: C-to-V effects

### 8.2.1 Methods

#### Speakers

Four male speakers participated in this experiment. They are MN, SO, KO and SK. MN is originally from Kagoshima, Kyushu in the South of Japan.[3] SO is originally from Kanagawa, a prefecture to the West of Tokyo. KO has lived in Tokyo since the age of five. SK is originally from Tokyo. They all speak Standard Japanese. KO did the recording for Experiment 3 as well. All of the speakers have had some experience of living abroad (in the United States and/or in Great Britain). At the time of the recording they were living in Edinburgh.

#### Materials

The VCaCV sequences with the consonantal contexts of /p, t, k/ and the vocalic contexts of /i, a, u/ were embedded in the sentences below. The data from this experiment will also serve as a reference data for a later experiment on interlanguage where the production of schwa by Japanese speakers of English will be compared with the native speakers' production of schwa. The vowel /a/ was selected as the target vowel for contextual assimilation as it is the most likely candidate among the five vowels of Japanese to be transferred from the L1 to the L2 schwa system. Because of the allophonic variation of /t/ before the high vowels /i/ and /u/, the sequences /itati/ and /utatu/ are realized as [itaʧi] and [ɯtatsɯ] respectively. It should also be noted that the vowel [ɯ] in Japanese is compressed (See Footnote 2 in Chapter 5), and therefore has higher $F_2$ values.[4]

Of the three contextual vowels, the two high vowels /i/ and /u/ are prone to the vowel devoicing assimilation in Japanese. (See 5.5 (p 106) for a detailed discussion of vowel devoicing in Japanese.) Therefore, in some cases both the

[3]MN's reading of the material did not deviate from the Standard Japanese accent apart from the tonal pattern of the word *na'tte-iru* (to bear a fruit). He pronounced it as an unaccented word, but this word was not of relevance to the present study, and his recording was used.

[4]The $F_2$ values for the English vowels [u] are also quite high. The mean $F_2$ values obtained in Experiment 1 for the vowel [u] are 1648 Hz, 1671 Hz and 1674 Hz for AH, MB and DG.

preceding and the following vowels were deleted as [astats] in *a*s**u** *tats*u *yotei desu.* Of the 60 contextual /i/'s, MN devoiced 17 (28 %) tokens, SO, 17 (28 %) and KO, 7 (12 %) tokens. SK voiced all the /i/ tokens. All the devoiced /i/'s are $V_2$ in VtaʧV or VkakV sequences. Of the 60 /u/'s, 25 (42 %), 40 (67 %), 33 (55 %) and 10 (17 %) tokens were devoiced by MN, SO, KO and SK respectively. Of these 14 tokens are $V_1$ and 11 are $V_2$ for MN, 22 tokens are $V_1$ and 11 are $V_2$ for KO, 4 tokens are $V_1$ and 6 are $V_2$ for SK. SO devoiced all the /u/'s in VtatsV and VkakV sequences. The effects of the vowel devoicing on V-to-V coarticulation in Japanese will be discussed below.

Prosodic effects could not be controlled in these materials. In some cases, the vowel /a/ was accented as in *pa'pirusu* (HLLL) while in other cases it was unaccented as in *tatakai* (LHHH). Accentual phrasing also seems to have varied from speaker to speaker, or even from token to token. The effects of accent or pitch on the vowel quality of Japanese were observed to be minimal, particularly in $F_2$, in Experiment 3. It was also observed that the presence of accent did not significantly affect the extent of V-to-V coarticulation in Japanese (Experiment 4a). Therefore, it was considered that the presence or absence of accent on the vowel /a/ would not affect the results of the present study in a seriously significant way though different accentual phrasing may have affected the results by placing a pause within the VCaCV sequences.

1. ipapi: Kawa-n**i** **pa'pi**rus-no ha-ga shige't-te-iru. (Papyrus reeds are growing in the river.)

2. apapa: Kore-g**a** **pa'pa**-no nigaoe desu. (This is Papa's portrait.)

3. upapu: As**u** **pa'pu**a nyuu ginia-ni iku. (We are going to Papua New Guinea tomorrow.)

4. itaʧi: Asoko-n**i** **tachi**noki-no huda-ga ta't-te-iru. (There is a notice board standing there telling us to leave the premises.)

5. atata: Kore-g**a** **tata**kai-no o'ngaku desu. (This is a war music.)

6. utatsu: Ka're-wa nihon-ni as**u** **ta'tsu** yotei desu. (He is leaving for Japan tomorrow.)

7. ikaki: Ano eda-**ni** **kaki**-no mi-ga na't-te-iru. (Persimmons are ripe on that branch of the tree.)

8. akaka: Kore-g**a** **kaka**shi-no e desu. (This is a picture of a scarecrow.)

9. ukaku: Henji'-o as**u** **ka'ku**-you-ni tsutae-te-kudasa'i. (Please tell him to answer the letter tomorrow.)

Each sentence was repeated 10 times by each speaker in a randomized order.

#### Recordings

Recording was done in a sound treated room in the Language Laboratory of the University of Edinburgh.

#### Analyses

The sentences for analysis were sampled at 16 kHz into a UNIX SUN workstation with WAVES speech analysis facilities. Formant values were obtained by running the FORMANT program for LPC analysis with 25 ms cos**4 window moving in 5 ms steps. Statistical analyses were performed with the vowel midpoint frequency values. (See 3.2.1: p 42 for a more detailed description of the analysis method.)

In order to obtain the vowel power, the program PWR was used. In this program, the raw power is computed by summing the squares of the sampled data values and dividing by the number of points in the frame. The (base 10) log of the raw power is then calculated and multiplied by 10 to convert to dB in order to compare the differences in power between vowels within an utterance.[5]

### 8.2.2 Results and discussion

Two-way ANOVAs were performed with consonantal and vocalic contexts as grouping factors for each speaker. Table 8.1 shows the level of significance for the two main effects and the group interaction on the first and second formants of the vowel /a/ for each speaker.

[5]For more discussion on the power, sound pressure, decibel, intensity and loudness, see Borden & Harris (1980:39-42) and Ladefoged (1962).

| | p-value | | | | | |
|---|---|---|---|---|---|---|
| | $F_1$ | | | $F_2$ | | |
| Subj. | $C$ | $V$ | $C \times V$ | $C$ | $V$ | $C \times V$ |
| MN | 0.0000 | 0.0001 | 0.0596 | 0.0000 | 0.0000 | 0.1217 |
| SO | 0.0306 | 0.0002 | 0.2644 | 0.0000 | 0.0000 | 0.0713 |
| KO | 0.0007 | 0.3222 | 0.0026 | 0.0000 | 0.0000 | 0.0000 |
| SK | 0.0023 | 0.0003 | 0.9924 | 0.0000 | 0.0000 | 0.0003 |

**Table 8.1.** The level of significance of the results of ANOVAs. The table shows the main effects of consonant and vowel contexts and their interaction on the first and second formants of the vowel /a/ for the four speakers.

The main effect of consonantal context was significant for both $F_1$ and $F_2$ for all of the speakers. The main effect of vocalic context was significant for both $F_1$ and $F_2$ for MN, SO and SK. KO did not show any significant effects of vocalic context in $F_1$. KO showed a significant interaction between the consonant and vowel contexts in both $F_1$ and $F_2$. SK showed a significant interaction in $F_2$.

Table 8.2 shows the matrices of the mean $F_1$ and $F_2$ values of the vowel /a/ measured at the vowel midpoint in the 3 consonant $\times$ 3 vowel contexts for each speaker. The rightmost value on the bottom row is the grand mean for each matrix.

The $F_1$ values of the vowel /a/ were generally higher in the context of the labial /p/ than in those of alveolar or velar consonants. A similar observation was made for the English schwa (see page 44).

The mean $F_1$ values were higher in the contexts of /i/ and /u/ than in the context of /a/ for MN, SO and SK. This is contrary to what one would expect as the vowels /i/ and /u/ are [+high] and their effects are generally considered to be the lowering of $F_1$ values. It seems as though the three subjects are overshooting the $F_1$ target of the vowel /a/ to compensate for the effects of the vocalic contexts.[6] KO did not show significant effects of the transconsonantal vowels on

[6]Broad & Fertig (1970) have observed a pronounced tendency for the first formant trajectories of the American vowel /I/ to overshoot as a function of the initial consonants in CVC syllables. That is, the initial consonants pulled up the $F_1$ values of the vowel /I/. This is contrary to what one would expect as the consonantal contexts are generally considered to lower

| $F_1$ | | | | | $F_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| MN | p | t | k | | MN | p | t | k | |
| i | 728 | 627 | 648 | 666 | i | 1326 | 1498 | 1396 | 1409 |
| a | 746 | 571 | 508 | 611 | a | 1153 | 1397 | 1296 | 1302 |
| u | 665 | 665 | 605 | 643 | u | 1189 | 1394 | 1258 | 1283 |
| | 721 | 620 | 592 | 640 | | 1233 | 1431 | 1317 | 1334 |
| SO | p | t | k | | SO | p | t | k | |
| i | 620 | 567 | 654 | 606 | i | 1264 | 1634 | 1397 | 1433 |
| a | 555 | 471 | 549 | 516 | a | 1119 | 1558 | 1377 | 1351 |
| u | 516 | 523 | 518 | 519 | u | 1155 | 1543 | 1318 | 1339 |
| | 562 | 520 | 563 | 545 | | 1180 | 1579 | 1362 | 1374 |
| KO | p | t | k | | KO | p | t | k | |
| i | 684 | 666 | 616 | 655 | i | 1367 | 1570 | 1607 | 1515 |
| a | 663 | 614 | 652 | 643 | a | 1203 | 1495 | 1466 | 1388 |
| u | 658 | 626 | 645 | 643 | u | 1256 | 1416 | 1356 | 1342 |
| | 668 | 635 | 638 | 647 | | 1275 | 1494 | 1476 | 1415 |
| SK | p | t | k | | SK | p | t | k | |
| i | 700 | 673 | 636 | 669 | i | 1439 | 1621 | 1563 | 1548 |
| a | 661 | 651 | 617 | 644 | a | 1162 | 1507 | 1430 | 1366 |
| u | 720 | 698 | 672 | 698 | u | 1308 | 1449 | 1335 | 1364 |
| | 694 | 674 | 641 | 670 | | 1293 | 1526 | 1438 | 1422 |

**Table 8.2.** The mean $F_1$ and $F_2$ values of the vowel /a/ measured at the vowel midpoint in the 3 consonant /p, t, k/ $\times$ 3 vowel /i, a, u/ contexts for each speaker. The rightmost column of each matrix shows the mean formant frequencies for the vocalic contexts. The bottom row shows the mean formant frequencies for the consonantal contexts. The rightmost value on the bottom row is the grand mean for each matrix.

the first formant value.

The mean $F_2$ values increased according to context, u < a < i, for all of the subjects. Differences in the mean $F_2$ values as a function of vocalic contexts are 126 Hz, 94 Hz, 173 Hz and 184 Hz for MN, SO, KO and SK respectively. The mean $F_2$ values increased according to context, p < k < t. The data from Stevens & House (1963) and Recasens (1986) show that the $F_2$ values of front vowels increase according to context, p < t < k, while those of back vowels increase according to context, p < k < t. From this it seems that the Japanese vowel /a/ is behaving as a [+back] vowel. Keating & Lahiri (1993) suggest that the velar /k/ is unspecified in [Backness]. How a vowel interacts with its adjacent velar could be a measure of the [Backness] of the vowel. Among the 'a'-type vowels, the vowel /æ/ in English is fronted and has higher $F_2$ values in the context of /k/ and thus, its $F_2$ values increase according to contexts, p < t < k. In comparison, the Japanese vowel /a/ is relatively back and its $F_2$ values increase according to contexts, p < k < t. Differences in the mean $F_2$ values as a function of consonantal contexts are 198 Hz, 399 Hz, 219 Hz and 233 Hz for MN, SO, KO and SK at the vowel midpoint.

KO showed an interaction between the consonantal and vocalic contexts in both $F_1$ and $F_2$. In the context of the front vowel /i/, his mean $F_2$ values of the vowel /a/ increased according to context, p < t < k, while in the contexts of the back vowels /a/ and /u/, the $F_2$ values increased according to context, p < k < t. That is, KO's vowel /a/ behaved like a front vowel in the context of front vowels in its relation to the contextual consonants. As shown in Table 8.6, KO shows the greatest effects of both vocalic and consonantal contexts in $F_2$ among the four speakers.

Comparing the formant values of the vowel /a/ obtained in this experiment with those obtained in Experiment 3, the $F_1$ values were higher and $F_2$ values were lower in Experiment 3 where the consonantal context was limited to /b/ (bVbV nonsense words). KO who participated in both experiments had the mean $F_1/F_2$ values of 688/1292 Hz in Experiment 3 with the consonantal context of

the $F_1$ value of a vowel. (Stevens & House 1963) On the other hand, more systematic effects of the contextual vowels were observed on the $F_1$ value of the Japanese vowel /e/ in Experiment 4a.

/b/ while he had the mean $F_1/F_2$ values of 647/1415 Hz across 3 consonant $\times$ 3 vowels. The labial consonants have been observed to lower the $F_2$ values of both front and back vowels. On the other hand, $F_1$ values are generally higher in the labial contexts. The effects of consonantal context seem to be very systematic.

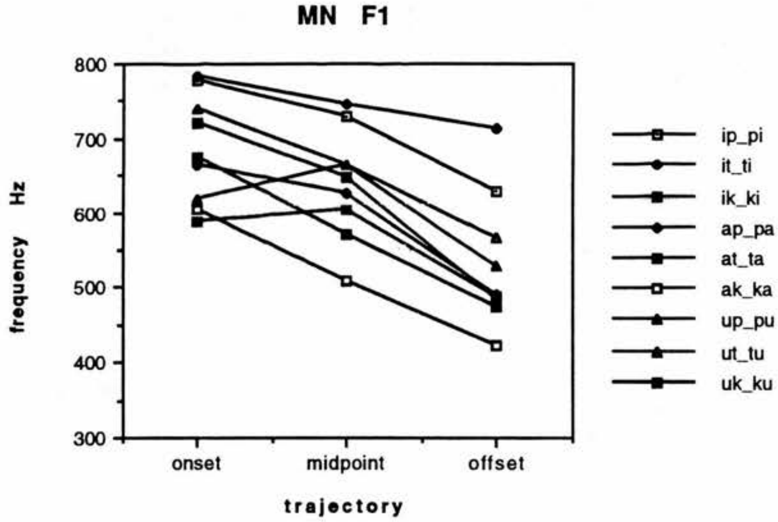
Figures 8.1 through 8.4 show the mean $F_1$ and $F_2$ trajectories from the onset to the offset of the vowel /a/ as a function of contexts for each speaker.[7] The first formant trajectories tend to decline from the onset to the offset for all of the speakers. A similar trend was observed in Experiment 4 with the VsekV sequences.[8] The ranges in frequency from the minimum to the maximum value of $F_1$ measured at the vowel midpoint are 405-993 (488) Hz, 366-861 (495) Hz, 533-754 (219) Hz and 505-756 (251) Hz for MN, SO, KO and SK respectively. The ranges are about 500 Hz for MN and SO while they are smaller for KO and SK. As the $F_1$ ranges for the English /ə/ were 148-395 (247) Hz, 226-480 (254) Hz and 203-391 (188) Hz for AH, MB and DG, it seems that schwa is less variable in $F_1$ than the Japanese vowel /a/. The scatterplots of individual tokens of the Japanese vowels produced in prose reading (Keating & Huffman 1984:Fig.4-10) also show a great spread in $F_1$ for the vowel /a/ for about half of the speakers. As the vowel /a/ in Japanese is far apart from the other four vowels in $F_1$ in the acoustic $F_1/F_2$ space, there seems to be much room for spreading. However, the mean $F_1$ values are rather stable across speakers. They are 640 Hz, 545 Hz, 647 Hz and 670 Hz for MN, SO, KO and SK. The mean $F_1$ values of the vowel /a/ for the 7 speakers in Keating & Huffman (1984) are 614 Hz, 605 Hz, 665 Hz, 656 Hz, 572 Hz,[9] 656 Hz and 647 Hz. This suggests that though a large spread in formant frequency as a function of contexts is allowed, there is a preferred 'centre of gravity' value for a particular vowel.

The ranges in $F_2$ are 1034-1555 (522) Hz, 1022-1698 (676) Hz, 1174-1659 (485) Hz and 899-1756 (857) Hz for MN, SO, KO and SK. These ranges are quite wide

[7]Each data point is the mean of 10 tokens. In some cases, there were less observations per mean due to formant tracking error. The smallest observation per mean is 5.

[8]The lowering of $F_1$ value at the vowel offset may be interpreted as a strong anticipatory effect of a consonant on the preceding vowel as consonants have been observed to lower the value of the adjacent vowels.

[9]This speaker has a rather crowded use of the vowel space with extensive overlap.



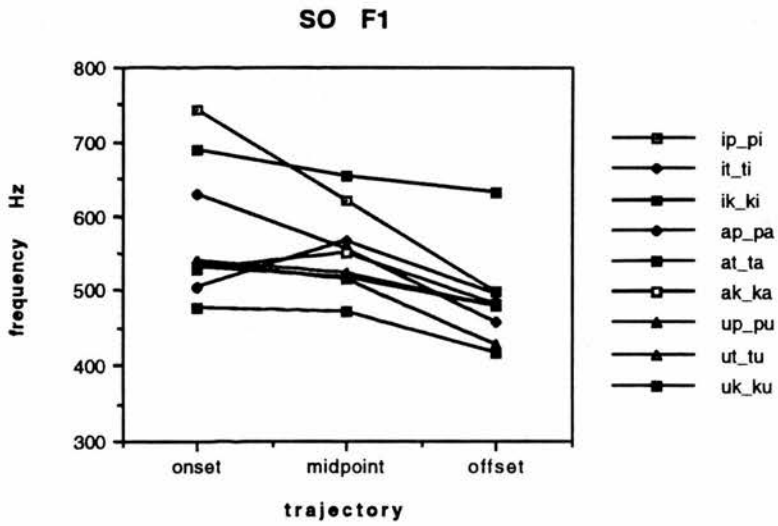




**Figure 8.1.** The mean $F_1$ trajectories as a function of contexts across the onset, midpoint and the offset of the vowel /a/ in Japanese for MN and SO.

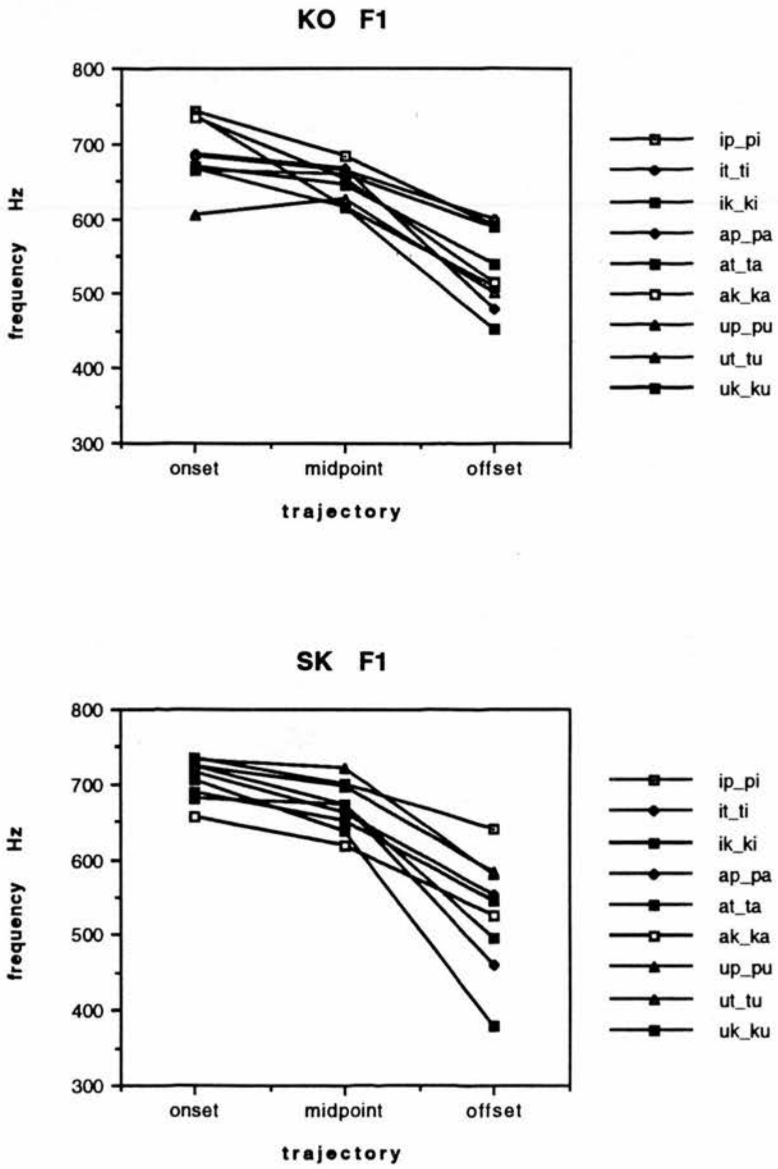



**Figure 8.2.** The mean $F_1$ trajectories as a function of contexts across the onset, midpoint and the offset of the vowel /a/ in Japanese for KO and SK.

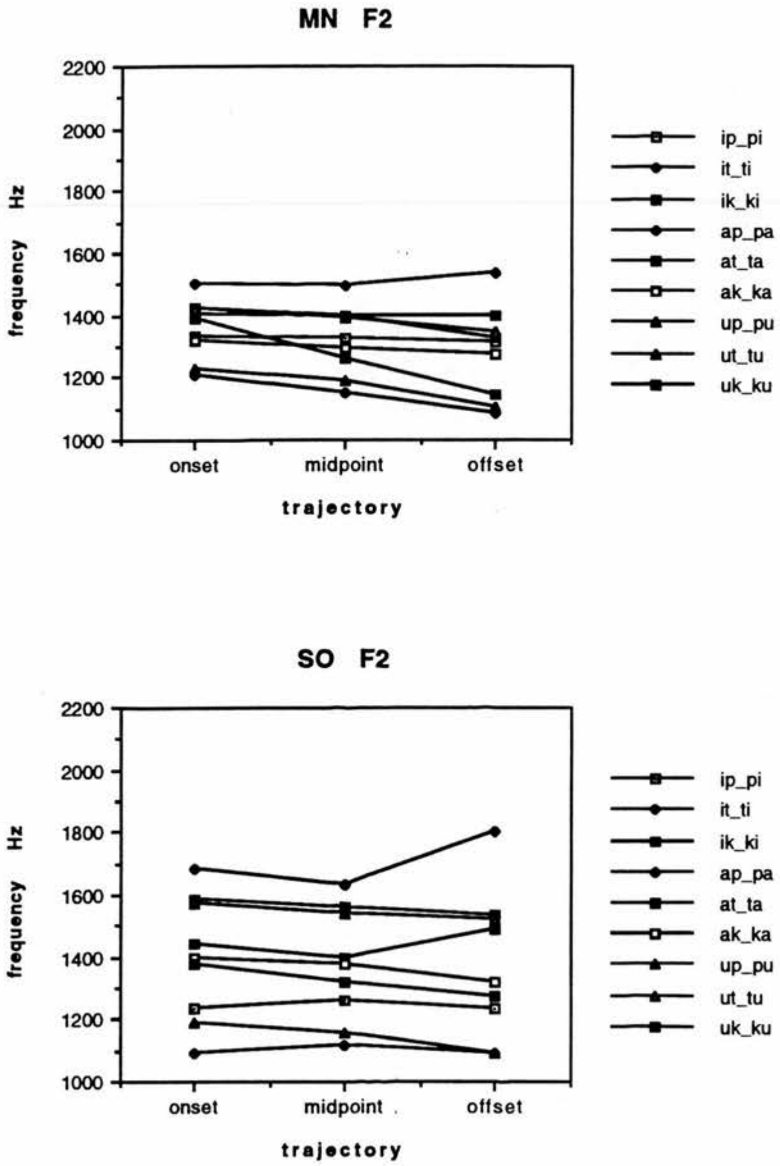
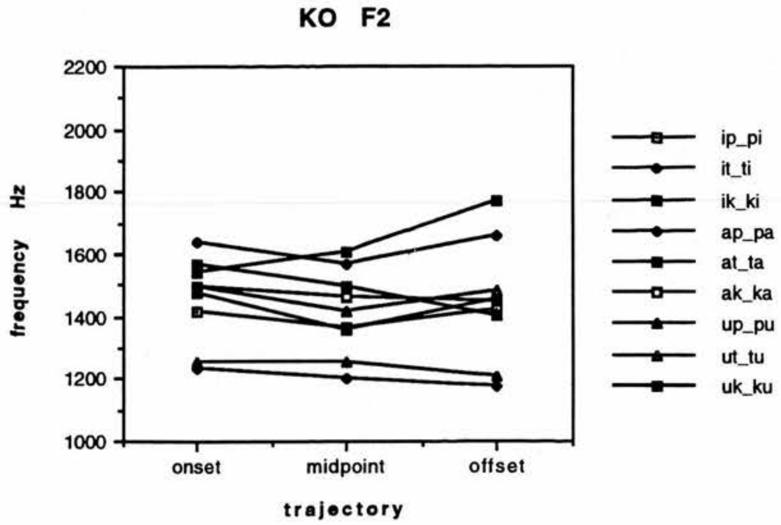



**Figure 8.3.** The mean $F_2$ trajectories as a function of contexts across the onset, midpoint and the offset of the vowel /a/ in Japanese for MN and SO.



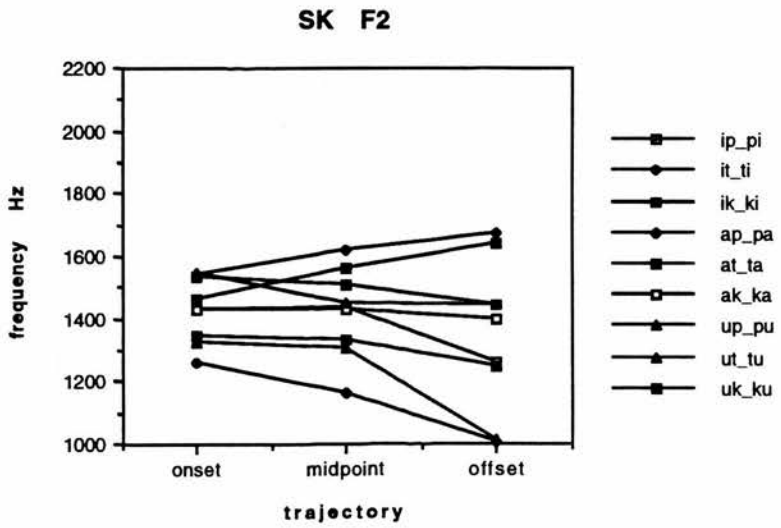




**Figure 8.4.** The mean $F_2$ trajectories as a function of contexts across the onset, midpoint and the offset of the vowel /a/ in Japanese for KO and SK.

as well.[10] SK shows the largest spread of 857 Hz. The trajectories tend to diverge towards the offset. In this respect the $F_2$ trajectories of the Japanese vowel /a/ are similar to schwa trajectories observed in Figure 3.6. In Japanese, however, differences in $F_2$ as a function of vocalic contexts, front vs. back vowels, are particularly great at the vowel offset, suggesting that anticipatory V-to-V effects may be greater than carryover effects for these speakers. For the English /ə/, the V-to-V effects were stronger at the offset than at the onset mainly in the context of /k/. (See Table 3.6.) In the present experiment, this effect was also observed in the contexts of /p/ and /t/, though to a lesser degree than in the context of /k/. Table 8.3 shows the differences in the mean $F_2$ values of the vowel /a/ as a function of the contextual vowels /i/ and /u/ at different points in the trajectory for different consonantal contexts. Note how the differences increase towards the offset of the vowel. This trend was also observed in Experiment 4.[11] The mean $F_2$ values at the onset were also generally higher than the mean $F_2$ values at the offset within each consonantal context. The differences of 99.3 Hz, 28.3 Hz and 23.1 Hz were observed in the labial, alveolar and velar context respectively. These asymmetries in the consonant loci at the onset and offset of the vowel may be explained by aerodynamic reasons (see page 52).

Table 8.4 shows the differences in the mean $F_2$ values of the vowel /a/ as a function of consonantal contexts in fixed vowel contexts. The C-to-V effects seem to be generally stronger at the offset for Japanese. This trend was particularly strong for SK. Similarly, for the English /ə/, anticipatory effects of the consonantal context were systematically stronger than carryover effects. The asterisks in Tables 8.3 and 8.4 indicate that the difference has reached the statistically significant level ($p < 0.05$) as a result of the post hoc scheffe tests for the interaction between vocalic context, consonantal context and the points of measurement for each speaker.

The extent of variability as a function of contexts was quite remarkable in both $F_1$ and $F_2$ for the vowel /a/ of Japanese. Though the $F_2$ trajectories of

[10] Figures 8.1 and 8.2 are scaled from the minimum of 1000 Hz to the maximum of 2200 Hz. This is the same scale as the one used in Figure 3.6, for the schwa trajectories for the purpose of comparison. Note how the trajectories of the Japanese vowel /a/ occupy the lower part of the graph while the trajectories of the English schwa occupy most part of the graph.

[11] See Figure 7.3.

| subject | context | onset | midpoint | offset |
|---|---|---|---|---|
| MN | Vp_pV | *105.6 | 137.6 | *209.3 |
| | Vt_tV | 81.5 | 103.9 | *193.3 |
| | Vk_kV | 9.7 | 137.9 | *255.6 |
| SO | Vp_pV | 46.4 | 109.4 | 140.2 |
| | Vt_tV | 112.9 | 90.5 | *280.2 |
| | Vk_kV | 63.5 | 79.0 | *216.8 |
| KO | Vp_pV | *165.6 | *111.1 | *211.7 |
| | Vt_tV | *141.2 | *154.4 | *174.8 |
| | Vk_kV | 67.3 | *251.0 | *312.7 |
| SK | Vp_pV | 101.8 | 131.5 | *250.4 |
| | Vt_tV | -6.5 | 172.3 | *228.2 |
| | Vk_kV | 119.9 | *227.9 | *392.1 |

**Table 8.3.** The differences in the mean $F_2$ values as a function of the contextual vowels /i/ and /u/ (/i/ - /u/) at the different points in the trajectory for different consonantal contexts. The asterisk means that the difference has reached the statistically significant level ($p < 0.05$).

| subject | context | onset | midpoint | offset |
|---|---|---|---|---|
| MN | iC_Ci | *171.7 | *171.3 | *225.1 |
| | aC_Ca | *213.4 | *244.3 | *245.2 |
| | uC_Cu | *195.8 | *205.0 | *240.7 |
| SO | iC_Ci | *450.4 | *369.4 | *567.1 |
| | aC_Ca | *494.3 | *439.0 | *443.3 |
| | uC_Cu | *383.9 | *388.3 | *427.1 |
| KO | iC_Ci | *218.2 | *203.4 | *238.3 |
| | aC_Ca | *328.0 | *291.9 | *223.7 |
| | uC_Cu | *242.6 | *160.1 | *275.2 |
| SK | iC_Ci | 111.6 | 182.3 | *407.5 |
| | aC_Ca | 272.0 | *345.0 | *432.2 |
| | uC_Cu | *219.9 | 141.5 | *429.7 |

**Table 8.4.** The differences in the mean $F_2$ values as a function of the contextual consonants /t/ and /p/ (/t/ - /p/) at the different points in the trajectory for different vocalic contexts. The asterisk means that the difference has reached the statistically significant level ($p < 0.05$).

the vowel /a/ do not range as widely as those of /ə/, these trajectories are quite diverse even at the midpoint. In $F_1$ as well, the trajectories cover a wide range at the midpoint for SO and KO, while for MN and SK, the trajectories from the onset to the offset seem to pass through some sort of target range. The mean durations for these trajectories are 46 ms, 42 ms, 58 ms, and 42 ms for MN, SO, KO and SK.

### 8.2.3 Comparison with schwa

Multiple regression analyses (BMDP 1R) were performed to see the predictability of the $F_1$ and $F_2$ values of the English schwa and the Japanese vowel /a/ as a function of consonantal and vocalic contexts. Since many of the contextual vowels were devoiced in Japanese and formant frequencies were not obtained, arbitrary scores of 1 to 3 were given for the vowels /i, a, u/.[12] Similarly, arbitrary scores of 1 to 3 were given for the consonants /p, t, k/. Tables 8.5 and 8.6 summarize the results of the regression analyses. Stepwise multiple regression analyses (BMDP 2R) were also performed with the same variables. The asterisks in Tables 8.5 and 8.6 show that the variables were entered into the linear regression equation.

Table 8.5 shows the results of the regression analyses for the English /ə/ spoken by native speakers. In $F_1$, significant effects of the consonantal context was observed only for MB. No significant effects of the vocalic context was observed in $F_1$. In $F_2$, both consonant and vowel contexts showed significant effects for all the speakers.

Table 8.6 shows the results of the regression analyses for the Japanese vowel /a/. Compared to schwa, there is less systematicity across speakers. The difference between schwa and the Japanese vowel /a/ is clearer in $F_2$. For schwa, all the subjects had $r^2$ values of over 0.6000 for the combination of C + V, while for the vowel /a/, only one out of four subjects had the $r^2$ value of over 0.6000. That is, more than 60 % of the total variance of schwa may be accounted for by the combination of the consonantal and vocalic contexts. In Japanese, KO

[12]The regression analyses were rerun for schwa with the arbitrary scores of 1 to 3 for the vocalic contexts /I, æ, u/ instead of using the midpoint $F_1/F_2$ values in order to make the comparison with the Japanese data possible.

| | | AH | MB | DG |
|---|---|---|---|---|
| | | $R^2$ | $R^2$ | $R^2$ |
| $F_1$ | C | 0.0349 | *0.1686 | 0.0003 |
| | V | 0.0046 | 0.0042 | 0.0090 |
| | C + V | 0.0395 | 0.1719 | 0.0095 |
| $F_2$ | C | *0.4473 | *0.6680 | *0.6584 |
| | V | *0.1707 | *0.0682 | *0.0989 |
| | C + V | *0.6170 | *0.7442 | *0.7330 |

**Table 8.5.** The results of the multiple regression analyses for the $F_1$ and $F_2$ variability at the midpoint of the vowel /ə/. The independent variables are consonantal context (C) and the vocalic context (V). For consonantal contexts, arbitrary scores of 1 to 3 were given for /p, t, k/, and for vocalic contexts, scores of 1 to 3 were given for /ɪ, æ, u/.

| | | MN | SO | KO | SK |
|---|---|---|---|---|---|
| | | $R^2$ | $R^2$ | $R^2$ | $R^2$ |
| $F_1$ | C | *0.2315 | 0.0007 | *0.0949 | *0.1197 |
| | V | 0.0194 | *0.1392 | 0.0147 | 0.0404 |
| | C + V | 0.2400 | 0.1426 | 0.1097 | 0.1553 |
| $F_2$ | C | 0.0244 | *0.1578 | *0.3791 | *0.1548 |
| | V | *0.1536 | *0.0842 | *0.2775 | *0.2162 |
| | C + V | 0.1920 | *0.2642 | *0.6567 | *0.3585 |

**Table 8.6.** The results of the multiple regression analyses for the $F_1$ and $F_2$ variability at the midpoint of the vowel /a/. The independent variables are consonantal context (C) and the vocalic context (V). For consonantal contexts, arbitrary scores of 1 to 3 were given for /p, t, k/, and for vocalic contexts, scores of 1 to 3 were given for /i, a, u/.

was the only subject who showed such strong effects of the context. For schwa, all the subjects showed stronger effects of the consonantal context than those of the vocalic context. For the vowel /a/ of Japanese, two subjects showed greater effects of the consonantal context, while the other two showed stronger effects of the vocalic context. In general, strong C-to-V effects that were observed on schwa could not be observed for the vowel /a/ in Japanese.[13] Also, while the consonant /t/ blocked the V-to-V effects on schwa, the consonant /t/ or /ts/ or /ʧ/ does not seem to block such effects for the Japanese vowel /a/.

### 8.2.4 Effect of vowel devoicing on V-to-V coarticulation

SO devoiced all the contextual /u/'s in VtatV and VkakV sequences. A separate set of ANOVAs were performed for his 40 /a/ tokens in VtatV and VkakV sequences with the vocalic contexts of /i/ and /u/, of which 17 /i/'s in the second V and all the tokens of /u/ (both the preceding and the following /u/'s) were devoiced. The main effect of vocalic context was significant for both $F_1$ and $F_2$: $F(1,31) = 8.09$, $p = 0.0078$ for $F_1$ and $F(1,31) = 41.66$, $p < 0.0001$ for $F_2$. The $F_1$ values of the vowel /a/ were lower in the context of /u/ than /i/. This may be due to overshooting as the $F_1$ values of /u/ have been observed to be higher than those of /i/ in the previous studies. (See Tables 5.1 and 6.1.) The $F_2$ values of /a/ were higher in the context of /i/ than /u/. This is what one would expect as a result of the effects of the front vs. back vowels. Though the devoiced vowels are subsumed under the frication of /s/, /ts/ and /k/ and appear to be deleted on the spectrogram, they are still there affecting the formant values of the vowel /a/ at the midpoint. This seems to support the gestural overlap account of the vowel devoicing in Japanese. Figure 8.5 shows spectral patterns taken at the midpoint of the voiced vowel /u/ in *asu*, and at the end of /s/ in *asu* where /u/ is deleted. The peaks coincide around 1469.4 Hz for the two spectral patterns. The mean $F_2$ value of the vowel /u/ obtained by Keating & Huffman (1984) is 1419 Hz. The corresponding waveforms and spectrograms are also shown.

[13] Compare also Table 8.4 with Table 3.7 (p 62).

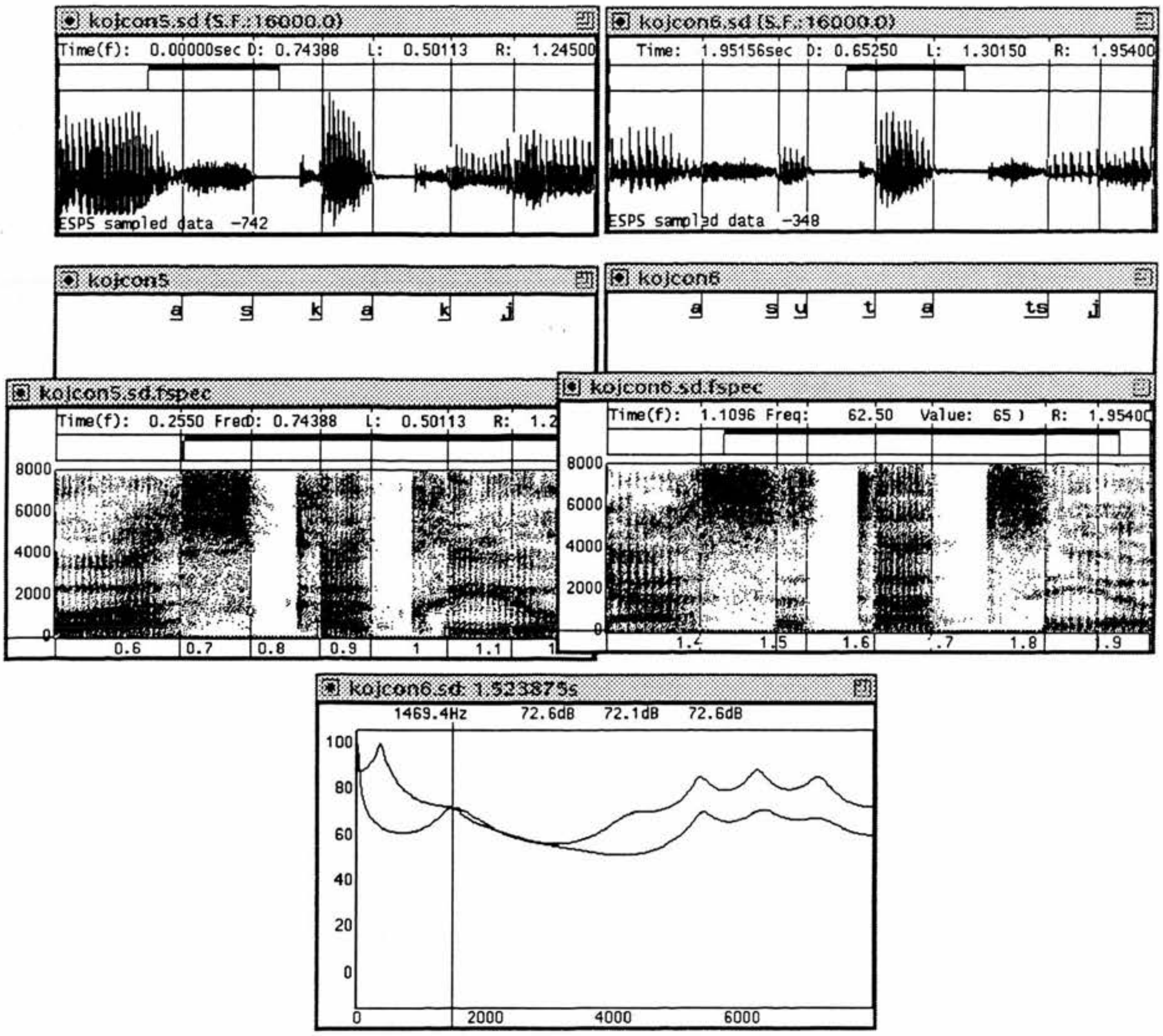



**Figure 8.5.** The spectral patterns taken at the midpoint of the voiced vowel /u/ in 'asu' and at the end of /s/ in the devoiced syllable /su/ of 'asu'

### 8.2.5 Rhythmic effect

For the aCaCa sequences, /apapa/, /atata/ and /akaka/, two identical morae occur consecutively in the words *pa'pa*, *tatakai* and *kakashi*. In these cases, the vowel in the first mora was usually shorter in duration[14] and smaller in power (Tables 8.7 and 8.8) regardless of the accent condition. That is, for the word *pa'pa*, the accent falls on the first mora, but the first vowel was usually weaker than the second vowel. The words *tatakai* and *kakashi* are unaccented, and in traditional tonal transcription are LHHH and LHH respectively. Differences in duration as a function of vowels ($V_1$ vs. $V_2$) were significant ($p < 0.001$) for all of the subjects. This suggests that some sort of rhythmic constraint is at work in favour of the alternation of weak and strong vowels. Han (1962b) also observed a similar constraint in the process of vowel devoicing. The onomatopoeic expression *pukupuku* where each vowel is in principle subject to devoicing yields the patterns [pu̥kupu̥ku] or [puku̥puku̥] in preference to other patterns where two adjacent vowels get devoiced.[15] Martin (1952) also suggests that devoicing of non-high vowels usually occurs in an initial syllable preceding an identical syllable, e.g., *ko̥koro*. The rhythmic constraint seems to happen where there is a succession of identical syllables or morae. Also in Experiment 4b, stronger V-to-V effects were observed where there was a succession of two identical vowels.

Table 8.8 shows the difference in amplitude between the two vowels expressed in dB. The negative value means that the first vowel is weaker in amplitude. The values are grouped according to the consonantal contexts. Note that in the context of /p/, the first vowel is accented. When the first vowel is accented, KO showed greater amplitude in the first vowel than in the second vowel while for

[14]The moraic durations were not measured as sometimes there was a pause between the first V and the following CVCV in the VCVCV sequence, and the duration of the first CV could not be measured accurately.

[15]Kubozono (1987) also observed a similar rhythmic constraint in catathesis. A monotonously descending $F_0$ contour is blocked by 'metrical boost' when it encounters a syntactically right-branching structure which is marked in Japanese: e.g., [kowa'i [me'no ya'mai]] (Terrible eye disease). Even in a uniformly left-branching structure, where each successive accentual phrase is expected to have lower $F_0$ than its preceding accentual phrase, every other accentual phrase is realized at higher $F_0$ than the preceding one to break the monotonous descend by the 'principle of rhythmic alternation'. According to Kubozono, catathesis does occur in such cases, but it is phonetically masked by 'metrical boost' and 'rhythmic boost'.

| | p | | t | | k | |
|---|---|---|---|---|---|---|
| | $V_1$ | $V_2$ | $V_1$ | $V_2$ | $V_1$ | $V_2$ |
| MN | 43 | 65 | 33 | 44 | 28 | 57 |
| SO | 34 | 62 | 36 | 38 | 25 | 62 |
| KO | 50 | 75 | 44 | 65 | 46 | 66 |
| SK | 34 | 78 | 33 | 55 | 27 | 62 |

**Table 8.7.** The mean duration of the first ($V_1$) and the second ($V_2$) of the identical CVCV sequences in different consonantal contexts.

| | p | t | k |
|---|---|---|---|
| MN | -1.24834 | -5.77639 | -5.82290 |
| SO | -1.85070 | -0.98736 | -6.11607 |
| KO | 4.11185 | -5.54054 | -5.56379 |
| SK | -5.90354 | -5.15303 | -8.74393 |

**Table 8.8.** The difference in dB between the amplitude of the first and the second vowel in the sequences *papa*, *tata* and *kaka*.

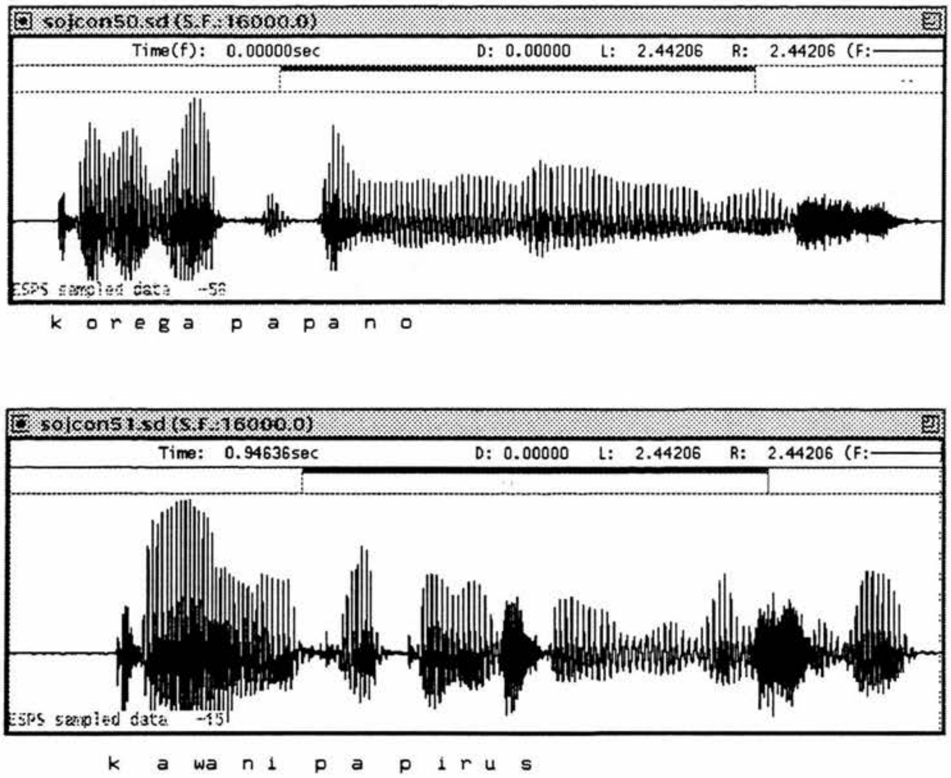



**Figure 8.6.** The waveform of the word 'pa'pa' and 'pa'pirus'. Note that the first vowel is weakened in 'pa'pa'.

unaccented words, he showed smaller amplitude in the first one. For MN, the first vowel was always weaker than the second vowel, but the difference in amplitude was smaller for the initially accented word. For these subjects, accent seems to override the rhythmic constraint discussed above. For SO and SK, the first vowel was always weaker than the second one. The contrast was particularly stronger for the word *kakashi*. The grouping factor of the consonantal context (papa, tata, kaka) was significant for all the subjects ($p < 0.01$). There is much interspeaker variability in behaviour as well as variability from one lexical item to another.

### 8.2.6 Summary of results

1. The effects of consonant as well as vowel context were significant for both the $F_1$ and $F_2$ values of the vowel /a/ in Japanese. (The only exception is the vocalic effect on $F_1$ for KO.) Unlike the English /ə/ where the effects of the adjacent consonants are very strong, the Japanese vowel /a/ did not show such strong effects of consonants. For two speakers the vocalic effects were stronger than the consonantal effects in $F_2$.

2. In $F_1$, the effects of the vocalic context were contrary to what one would expect. In the context of [+high] vowels where the $F_1$ values are low, the $F_1$ values of the target vowel /a/ were higher. It seems as though the subjects were overshooting to compensate for the effects of the contexts. In general, the $F_1$ values were higher in the context of the labial consonant /p/.

3. The effects of the vocalic context on $F_2$ were straightforward. That is, the vowels in the context of front vowels had higher $F_2$ values, while the vowels in the back vowel context had lower $F_2$ values.

4. The $F_2$ values of the vowel /a/ increased according to context, $p < k < t$. The vowel /a/ in Japanese seems to be behaving as a [+back] vowel.

5. The $F_1$ trajectories of the vowel /a/ were declining from the onset to the offset. The ranges in $F_1$ values were quite wide, from about 200 Hz to nearly 500 Hz.

6. The $F_2$ trajectories ranged quite widely as a function of contexts even at the vowel midpoint. The $F_2$ trajectories diverged towards the offset. There was a greater spread in $F_2$ values as a function of contexts at the offset, suggesting that anticipatory coarticulation may be stronger than carryover effects in Japanese.

7. The vowels that were deleted as a result of vowel devoicing process still affected the target vowel in the expected direction. That is, the vowels in the context of devoiced /u/'s had lower $F_2$ values than those in the context of /i/, some of which were also devoiced.

8. When there was a sequence of two identical morae in succession, the vowel in the first mora was usually shorter in duration and smaller in amplitude. Some sort of rhythmic constraint seems to be at work. For some speakers, the presence of accent interfered with this tendency, while for others, the rhythmic constraint seems to operate independent of the accent condition.

## 8.3 Conclusion

The C-to-V effects on the Japanese vowel /a/ were observed using the same VC_CV context used to look at C-to-V and V-to-V effects on schwa. Though significant C-to-V effects were observed for both $F_1$ and $F_2$, the magnitude of such effects was smaller on the Japanese /a/ compared to /ə/.

## 8.4 General conclusion

In Experiments 3 through 5, possible sources of vowel variation in Japanese have been explored. The effect of accent was of a particular interest. Accent was observed to raise the $F_1$ values in Experiment 3 and 4. It seems that a generally more open articulation characterizes accent. On the other hand, accent and pitch seem to have very little effect on the $F_2$ values. Where there was a significant difference as a function of accent, unaccented vowels had higher $F_2$ values. The magnitude of vowel variability was not affected by the presence of accent. For some speakers, accented vowels were more variable than unaccented ones and for others, the opposite was true. Vowel length showed significant effects on both $F_1$ and $F_2$ values. Phonemically short vowels tend to fill in the more central part of the vowel space. Whether this is a centralization, in other words a tendency for vowels to degenerate into a more neutral position, or a result of increased contextual assimilation is an open question. However, the results of the study on schwa (Chapters 3 & 4) suggest that the 'increased contextual assimilation' may be a more plausible account. Both the consonantal and vocalic contexts significantly affected the $F_1$ and $F_2$ values at the midpoint of the vowels /e/ and /a/ in Japanese (C-to-V effects were not studied for the vowel /e/). The C-to-V effects observed on Japanese vowel /a/ were smaller in magnitude compared to those observed on the English /ə/. On the other hand, strong V-to-V effects were observed on the Japanese vowels /e/ and /a/. The extent of V-to-V effects observed on the vowel /a/ was sometimes comparable to or even greater than that observed on /ə/ both in magnitude and across time depending on the conditions. The directionality of V-to-V effects and the segment beyond the affected segment seem to affect the overall strength of V-to-V coarticulation in Japanese. Anticipatory V-to-V effects were very much pronounced for the two vowels for every speaker.

# Part III

# INTERLANGUAGE

# Chapter 9

# Issues in Second Language Acquisition

In the present chapter, a brief review of some important issues in second language acquisition (SLA) studies will be presented as a background. The issues in SLA are diverse and complex with competing models and theories. Comprehensive coverage of such a vast area is beyond the scope of this dissertation. Nor is an account of the historical development of the field intended in the present study. Instead, some current issues in SLA that are relevant to the data observed in the present study will be discussed in detail. First of all, a definition of interlanguage will be given. Secondly, some important characteristics of interlanguage that seem to explain the data obtained in the present study will be introduced.

## 9.1 What is interlanguage?

The term interlanguage was coined by Selinker (1972) to describe cognitive processes through which second language (L2) learners evolve their own grammatical system of L2. Ellis (1994) defines interlanguage as follows,

> ...the internal system that a learner has constructed at a single point in time ('an interlanguage') and ... the series of interconnected systems that characterize the learner's progress over time ('interlanguage' or the 'interlanguage continuum'). (Ellis 1994:p350)

Ellis (1994) lists the following three questions as the main issues in interlanguage studies.

1. What processes are responsible for interlanguage construction?

2. What is the nature of the interlanguage continuum? In other words, how does interlanguage evolve over time?

3. why do most learners fail to achieve full target language competence?

The first two questions involve the issue of what sources of knowledge L2 learners have at the starting point of the development of interlanguage. Two important sources of the knowledge system of L2 are the first language (L1) and linguistic universals. The processes that may contribute to the construction and development of interlanguage are variability, hypothesis testing, implicit and explicit knowledge. They may affect the interlanguage both internally and externally. The last question may be related to 'fossilization' and 'merged' system. These concepts will be described below.

Some of the important assumptions behind the interlanguage studies may be listed as follows. I will briefly go through these assumptions.

- Interlanguage is a natural language and therefore systematic.

- It is dynamic and variable in nature.

- There are competing grammars or rules in interlanguage which give rise to variability and change.

- Interlanguage is a product of both conscious learning and communication strategies.

- Unlike L1 learners, L2 learners generally do not reach the same target level of competence as native speakers.

The most basic and almost axiomatic assumption in SLA studies is that interlanguage is a natural language. Because it is a natural language, it is under the same rules and constraints as languages in general. As there is a synchronic state and diachronic stages in a language, interlanguage has an internally consistent system at every single point in time, while it evolves over time at much faster rate than other languages do. Thus, interlanguage is variable and dynamic. Learners are constantly introduced to new forms, and they keep testing different hypotheses to accommodate the new forms into new functions in their interlanguage (Corder 1976). These form-function relationships may not be the same as those observed in the target language, and learners will have to go through a number of changes before they get to the correct relationship.

In the process, they use both conscious learning and communication strategies. Interlanguage is based on both 'explicit' (conscious, metalingual and/or learned) and 'implicit' (intuitive, subconscious and/or acquired) knowledge (Krashen 1981). The area of interlanguage phonology seems to be one area where there has been little conscious learning. Remembering the way I was taught English, I do not recollect any memory of being explicitly taught how to pronounce English sounds apart from the distinction of notorious /r/ and /l/ until I became an undergraduate student majoring in language studies. Having myself stepped into the area of phonetics, it seems that it is extremely difficult to have conscious awareness of the sound pattern of languages, whether it is the first language or the second language. It needs a certain amount of training to gain such awareness. At present not many language learners or even teachers are equipped with such awareness or have had opportunities for such training.

Lastly, it is often recognized that there is a limit to the development of interlanguage. L2 learners never reach the same level of competence as the speakers of the target language. This phenomenon is expressed by the term 'fossilization' by Selinker (1972). Flege & Hillenbrand (1984) describe the same phenomenon as a 'merged' system in interlanguage of phonology. Even casual observations are sufficient to show that foreign accent is extremely difficult to be disposed of even among fluent and advanced learners of L2.

In the following sections, linguistic universals, language transfer, variability and merged system will be discussed in detail. The concept of markedness will be introduced in the section under linguistic universals. The importance of transfer in interlanguage phonology will be discussed. Variability seems to be an important feature in describing interlanguage in phonology and phonetics. Lastly, the concept of 'merged' system will be introduced. Each of these concepts will provide theoretical backgrounds to the interpretation of data in subsequent chapters.

## 9.2 Linguistic universals

The concept of linguistic universals plays an important role in SLA research. There are two major approaches to linguistic universals. They are typological universal approach and Universal Grammar (UG) Theory. The basic assumption within this approach in SLA studies is that interlanguage is a natural language, and therefore it is under the same constraint of linguistic universals as the first language is. The study of typological universals was initiated by Greenberg (1966), and may be represented by the work of Comrie (1984), Croft (1990) and Hawkins (1983) among others. Typological universals are characterized by their implicational nature. For example, 'if a language has a noun before a demonstrative, then it has a noun before a relative clause' (Hawkins 1983). The issue of markedness in typological universals is closely related to SLA studies. Classical markedness defined by the Prague School of linguists is binary in nature. For example, in the pair of features 'a' and 'an', 'a' is unmarked whereas 'an' is marked. On the other hand, markedness viewed in typology studies is relative and implicational in nature. If a language that has property X also has property Y, then Y is unmarked in relation to X. That is, the existence of Y naturally follows from the existence of X in that language. Implicational hierarchies reflect degrees of markedness. The question related to the markedness issue in SLA are:

1. Is there any relationship between markedness hierarchy and the order of acquisition of grammatical features?
2. What effects does markedness have on learning difficulty?
3. How do typological factors in L1 and the target language (L2) interact in interlanguage?

Another important approach to linguistic universals is the theory of Universal Grammar as proposed by Chomsky (1976, 1981a, 1981b). Chomsky (1976:p29) defines UG as 'the system of principles, conditions, and rules that are elements or properties of all human languages.' These linguistic universals are considered to consist of principles and parameters. Principles are abstract properties of grammar that are generally observed in all human languages. Parameters are a set of finite options a particular language may draw on to determine the grammar of the language. Languages differ from one another in parameter setting. Parameters make it possible for languages to have variations. These variations are not random but systematic because they follow linguistic universals. In other words, parameters are already there in the UG, but they have to be set or 'switched on' by each specific language to be operative.

An example of a parameter is 'pro-drop'. Some languages, for example, Spanish and Italian, allow the deletion of subject pronouns while languages like English forbids it. The pro-drop parameter has only two settings, but there may be multiple settings for some parameters.

> *Is a girl.
> Es una muchacha.

Parameters like 'pro-drop' are interesting in that they involve a number of linguistic features that are considered to 'cluster'. For example, languages with null subjects do not have expletives, i.e., dummy 'it' and 'there' in English. They also permit variable word order. Each parameter thus encompasses a set of linguistic features that characterize the language.

> Viene la muchacha. (*Is coming the girl.)
> La muchacha viene. (The girl is coming.)

The concept of 'markedness' is defined differently in the UG framework. The degree of markedness depends on whether the feature is in the 'core' or in the 'periphery'. The core features of the language are those that are governed by UG, while peripheral features are those that are not. Core features are considered to be unmarked, while peripheral features are marked. The degree of markedness may also vary within the core as parameter settings may be ordered according to markedness. For example, Hyams (1983) considers pro-drop to be unmarked in relation to non-pro-drop. Markedness in the UG framework is different from that of the typological universal framework. In typological universals, markedness is external to learners. Learners must observe it and experience it in the input. On the other hand, markedness in UG is internal to learners. UG is an innate language knowledge and belongs to the language faculty of each individual. According to the theory, every human being is born with this faculty.

## 9.3 Transfer

> Transfer is the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired. (Odlin 1989:p27)

Views on transfer have undergone a considerable change. First it was advanced in the behaviourist framework of learning. According to their view, the 'habits' of L1 would be transferred to L2. The degree of difficulty was considered to depend on the extent to which the target language was similar to or different from L1. 'Positive' transfer would occur when the two languages pattern similarly. On the other hand, when the two languages are different, errors would result from 'negative' transfer. Positive transfer would facilitate learning while negative transfer interferes with it. Elaborate contrastive analyses of the native and target language were carried out to predict the pattern of L1 interference. However, the predictions of these studies were not empirically tested until the late 1960's. The results of subsequent error analysis studies cast doubts on the claims of the contrastive analysis hypothesis (CAH) as many of the predictions were not born out. This led to a rise of a minimalist position regarding L1 influence. Currently, however, the importance of transfer in SLA has been recognized by a number of researchers. This view is expressed by Odlin (1989:p4).

> Despite the counterarguments... there is a large and growing body of research that indicates that transfer is indeed a very important factor in second language acquisition.

Research has been focused on the question of under what conditions transfer is likely to occur. Some of the conditions that may affect transfer are language level, developmental factor and markedness. See Ellis (1994:Chapter 8) for more discussion on other factors, such as social factor, prototypicality, language distance and psychotypology.

Language level means different levels in language such as phonology, lexicon, syntax, semantics and discourse. Transfer seems to occur more readily in some levels than in others. For example, it is widely recognized that transfer is more pronounced at the level of the sound system than at the level of syntax. Flege (1981) introduced the concept of 'equivalence classification'. This is a tendency for learners to classify similar sounds in L1 and L2 into a single category. When the sounds in L1 and L2 are noticeably different, learners will develop a new target for this 'different' L2 phone. However, when the sounds in L1 and L2 are 'similar', learners fail to develop a new target and transfer occurs (Bohn & Flege 1992). This is because in the acquisition of the sound pattern of the first language, we have been trained to pay attention to certain distinctive features, and to ignore the rest of the vast acoustic mass that are present in order to allow for the wide range of allophonic variations within each phoneme. As a result, when we are faced with acoustically different phones such as those occurring at the beginning of the French and English words *taille* and *tie*, we seem to place them into the same category. Objectively listeners may detect the foreign accent in the first 30ms of /t/ bursts (Flege 1984), but speech perception seems to function categorically.

On the other hand, very young bilingual children who are exposed to two languages from the start may develop an 'enriched' system. They seem to learn the sounds occurring in the phonetic surface of their two languages independently as if each sound were a separate phoneme in an enriched pan-language system. For example, Leopold (1947) reports that a two-year-old exposed to English and German produced the word *ball* differently in those two languages. In English she said [bau] and in German [baɪ], thereby approximating the dark word-final /l/ of English and the light /l/ of German. Major (1977) also reports that his daughter started to differentiate the two languages English and Portuguese at the age of 1.9 to 2.1 and produced voiceless stops with aspiration when speaking

English, but without aspiration in Portuguese.

Another factor that might affect transfer is a developmental factor. For example, in the 'restructuring continuum' as suggested by Corder (1978), the starting point of L2 acquisition is the learner's L1, which is gradually replaced by the target language as acquisition proceeds. This view would suggest that transfer would be observed more at the earlier stage of the development than at a later stage. In some aspects of language, e.g., phonology, transfer seems to be a starting point as suggested by Wenk (1986). Wenk observed the acquisition of L2 English rhythm by French learners. Beginner learners simply transferred the French rhythm into their production of English while advanced learners showed standard English rhythm. Interestingly, intermediate learners showed a kind of hybrid rhythm system. Thus, he observed three different stages in the development of interlanguage starting from a simple transfer. Supporting evidence for this view has also been found by Major (1986). However, the picture is not so simple as Kellerman (1983) points out. Kellerman observed that some errors (involving pronominal copies in relative clause) that are clearly traceable to L1 influence emerged only at later stages of acquisition when the learner was sufficiently advanced and had better knowledge of the L2. Also transfer is never eliminated in some cases. This is also suggested in L2 phonology studies. As a result of 'equivalence classification' discussed above, learners will develop a 'merged' system (Flege & Hillenbrand 1984). In the above example of *taille* and *tie*, French learners of English (or English learners of French) will produce /t/ with the voice onset time (VOT) value that is intermediate in degree between those observed for the typical French /t/ and the typical English /t/. The results of long-term studies suggest that even after long exposure to L2, most L2 speakers never reach the target VOT value of the L2. The 'merged' system will be discussed in more detail below.

Transfer may also be affected by markedness. Learners seem more likely to transfer unmarked features in L1 than marked ones, particularly if the corresponding feature in L2 is marked (Zobl 1984).

## 9.4 Variability

Corder (1977) identifies both 'horizontal' and 'vertical' dimensions of interlanguage. The horizontal dimension refers to the interlanguage that a learner has at a particular point in time. The vertical dimension refers to the course through which interlanguage develops over time. The basic assumption of interlanguage is that it is systematic in nature. Though interlanguage is under constant change, it has an internal structure that is systematic and consistent within itself at every single point in time. The process of development from one stage to another is also considered to be ordered and systematic.

However, interlanguage is extremely variable at the same time. It is constantly exposed to new inputs. Learners are testing different hypotheses to accommodate the system into an internally consistent structure. That is, each interlanguage has some alternative rules for performing the same function. Learners shift from one rule to another in the course of hypothesis testing. Also, despite the striking similarity in the developmental pattern across different learners, there are variations in the way learners evolve their interlanguage. Interlanguage is an unstable system and it is susceptible to change from both internal and external stimuli. Interlanguage is variable in both horizontal and vertical dimensions though a great deal of variability may be characterized as systematic.

Researchers in the Chomskian tradition argue that the goal of SLA studies is to build a theory of L2 competence. In this sense, variability, which belongs to the domain of performance rather than competence, is of little interest to them. On the other hand, researchers in the functionalist tradition, that is, those who see that every different form has its own unique function, consider variability as an important part of their theory. Their approach is primarily sociolinguistic and/or psycholinguistic.

Three types of variability are proposed by Ellis (1992). They may be divided into systematic and non-systematic variability. Among the systematic variability, two major types are considered in SLA studies. They are situational and contextual variability.

Situational variability consists of a number of alternative linguistic forms that are dependent on extra-linguistic factors, for example, scene and participant (Brown & Fraser 1979). Scene includes setting, type of activity and subject matter in or on which the language is used. Participant covers certain characteristics of language users such as sex, age, social class and ethnicity. Situational variability is stylistic variability. Its research has its root in Labov's model (Labov 1970). Labov looked at certain sound variations in the speech of New Yorkers. He collected data from a range of speech styles from the 'vernacular' to the 'careful'. His subjects used socially prestigious sounds, such as /θ/ more in a situation where they had to pay more attention to their speech. In the casual style there were less instances of /θ/ than in the careful style. Tarone (1983) gives an excellent summary of SLA research on situational variability.

> ...the data we have examined...indicate that interlanguage does vary systematically with elicitation task and, further, that when a task elicits a relatively more careful style, that style may contain more TL (target language) forms and more prestige NL (native language) variants than the relatively more casual style elicited by other tasks (Tarone 1983).

The second type of variability that plays an important role in interlanguage is contextual variability. Language users vary their use of linguistic forms according to the linguistic environment. Dickerson (1975) found that the phonetic quality of specific phonemes produced by Japanese learners of English varied according to the phonetic contexts. In other words, there were systematic allophonic variations. For example, when /z/ was followed by a vowel, his subjects used the correct target form 100% of the time from the early stage of interlanguage development. On the other hand, when /z/ was followed by silence, they used other variants along with /z/.

The third type of variability is listed by Ellis (1992), that is, non-systematic variability or free variation. Though non-systematic variation is given little attention in SLA studies, it seems to be an important factor in triggering change in the development of interlanguage in phonology. An excellent account of the process involving non-systematic and systematic variability is given by Ellis (1992:p135).

> The learner will endeavour to maximise his linguistic resources by creating a system in which different forms serve different functions. The first stage consists of forms used in free variation, but subsequent stages involve the progressive sorting of forms into functional pigeon holes. It is likely that the first sorting will not establish form-function correlations that correspond to those of the target language. This may take several sortings and may never be entirely achieved. ...
>
> It is easy, now, to see why two types of variability arise in interlanguage. Non-systematic variation occurs when new forms are assimilated but have not yet been integrated into the learner's form-function system. Systematic variation occurs when the new forms have been accommodated by a restructuring of the existing form-function system to give the new forms their own meanings to perform. Situational variability is one aspect of this process.

## 9.5 'Merged' system

It has been reported in a number of studies (Caramazza *et al.* 1973; Williams 1980; Port & Mitleb 1981; Flege 1981; Flege & Hammond 1982; Flege & Hillenbrand 1984) that the voice onset time (VOT) of plosives achieved by L2 learners is somewhere between the values typically observed for L1 and L2 monolingual speakers. Similarly when L2 learners produce a contrast between a tense and a lax vowel, e.g., /u/ and /ʊ/ in English, where L1 lacks such a contrast, the acoustic distance between the members of such a pair is shorter in their interlanguage (Kondo 1989; Flege 1992). That is, though the two vowels produced by L2 speakers may be significantly different, they do not get as far apart in the $F_1/F_2$ vowel space as those produced by native speakers.

Flege (1980) proposes that a learner's phonetic control of the L2 sounds is the output of an interlanguage rather than the product of interference between L1 and L2 phonetic systems. The nature of this interlanguage as proposed by Flege (1981) is a 'merged' system in which the phonetic properties of L1 and L2 phones get merged. This interlanguage system evolves as a function of time and experience (Flege & Hillenbrand 1984). However, the extent to which the system approaches the target system seems to be limited. An example of this may be illustrated in the VOT studies cited above. Even after a considerable amount of exposure to L2, the speakers in these studies showed VOT values that were intermediate in degree between those observed for L1 and L2 monolinguals. Further, a native French speaker who had lived about 12.2 years in an English-speaking environment produced the /t/ in French words with VOT values that were half-way between the short-lag and long-lag values typically observed for French and English /t/'s. Thus, there seems to be influence back from L2 to L1.

Figure 9.1 illustrates the case of a 'merged' system. It shows the $F_1$ and $F_2$ values of the six British English vowels /ɪ, ɛ, æ, ɑ, ɔ, u/ (see Footnote 11 in Chapter 3 for the material description) produced by native and non-native (Japanese) speakers of English. Each data point is the mean across 10 repetitions (or less) by 6 subjects in each group.[1] In English there are two /a/-like vowels, /æ/ and /ɑ/, whereas in Japanese there is only one /a/. The acoustic distance between the vowels /æ/ and /ɑ/ produced by Japanese speakers was significantly apart only in $F_2$ ($F(1,90) = 17.90$, $p < 0.0001$). The acoustic distance between the two vowels in $F_1$ was not significant ($F(1,90) = 2.88$, $p = 0.0934$). Though Japanese subjects are showing some difference at least in $F_2$ (69 Hz), the difference is smaller compared to the native speakers' production (386 Hz). The native speakers showed significant difference between these vowels in both $F_1$ ($F(1,101) = 63.11$, $p < 0.0001$) and $F_2$ ($F(1,101) = 296.18$, $p < 0.0001$).

The results obtained by Beckman (1986) on perceptual cues to accent in English and Japanese illustrate an interesting example of an interlanguage system. She used three groups of subjects, Japanese, American bilinguals of Japanese and American monolinguals. She tested the effects of fundamental frequency, duration, amplitude and spectral coefficient pattern on the perception of accent in English and Japanese. Figure 9.2 taken from Beckman (1986:p184) shows the results. In interpreting the Japanese stimuli, the mean effect-on-accent scores for the fundamental frequency were higher for the Japanese subjects than for the American bilingual subjects, who in turn scored higher than did American monolinguals.

[1] Some of the tokens with spurious formant values were eliminated from the analyses as missing values. Thus, each data point may be smaller than $10 \times 6$ in some cases.

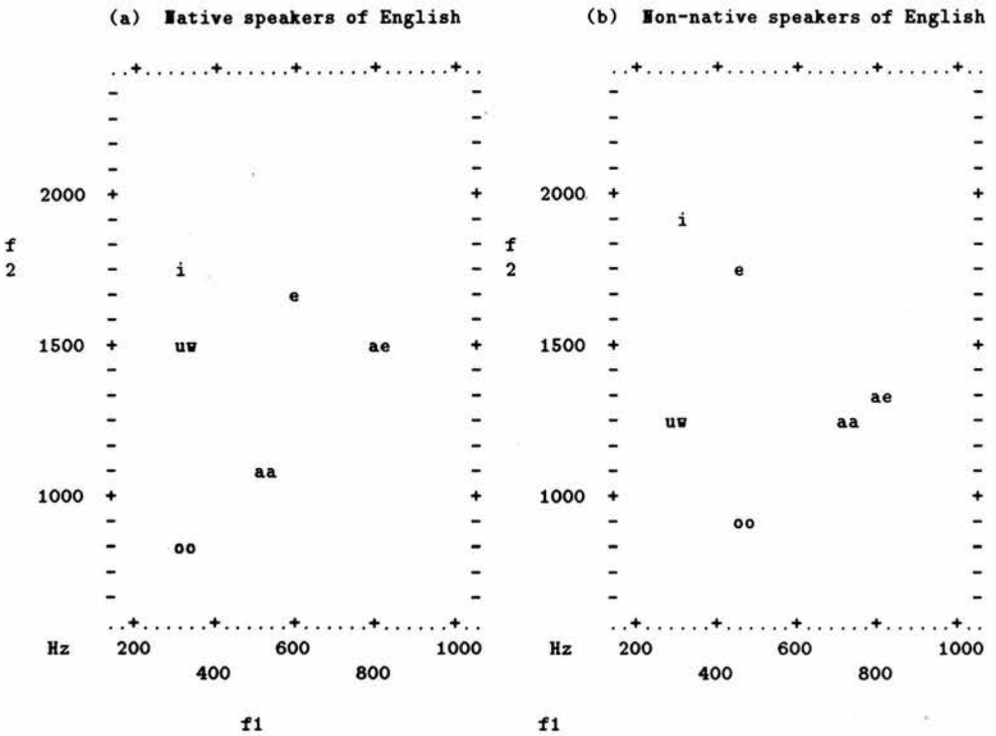



**Figure 9.1.** The mean $F_1$ and $F_2$ values of the 6 British English vowels /ɪ, ɛ, æ, ɑ, ɔ, u/ pronounced by native and non-native (Japanese) speakers of English. The symbols i, e, ae, aa, oo, uw represent the above 6 vowels respectively. Each data point represents the mean of 10 repetitions (or less) $\times$ 6 speakers.

The pattern of the scores for the English stimuli was even more interesting. The Japanese subjects used the $F_0$ to a much greater extent than the other cues. The American monolinguals, on the other hand, used all the four cues to a similar extent. The scores got higher in the order of the amplitude, spectral pattern, duration and fundamental frequency. The American bilinguals used the $F_0$ cue more extensively than did the American monolinguals in interpreting the English stimuli though they did not use it as much as did the Japanese subjects. The scores for the spectral pattern and duration made by the American bilinguals also fell somewhere in between the scores made by the Japanese and American monolingual subjects.

The fact that some of the Japanese subjects who participated in the experiment had been exposed to English for a number of years (some of them had lived in the United States over ten years or were married to Americans) suggests that it is more difficult for Japanese to learn to use cues other than pitch while it is relatively easy for American English speakers to make more extensive use of $F_0$. The fundamental frequency cue which is common to both stress-accent and non-stress accent languages may be universal and unmarked, therefore lending itself easily accessible to L2 learners.

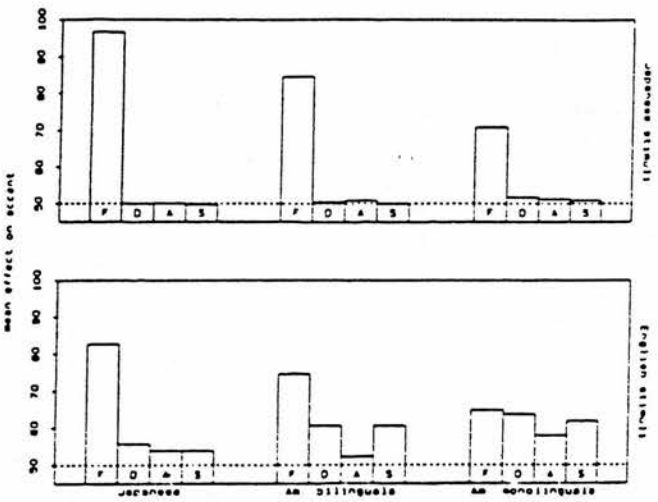



*Figure 7.1.* Mean effect on accent of the fundamental frequency (F), duration (D), amplitude (A), and spectral coefficient (S) patterns, averaged over responses to all tokens of Japanese stimuli (upper graph) and English stimuli (lower graph). Means are broken down by subject group.

**Figure 9.2.** Figure taken from Beckman (1986:p184).

# Chapter 10

# Interlanguage of Schwa 1

## 10.1 Introduction

In this chapter, schwa production by non-native (Japanese) speakers of English will be compared with those produced by native speakers of English. The key question in the present chapter is: will Japanese speakers of English acquire the important feature of schwa, i.e., its targetlessness in $F_2$?

Two groups of non-native subjects were studied in the present study to observe the nature of interlanguage of schwa at different stages. One group of subjects may be described as more native-like, and the other group as less native-like in their overall performance in the production of English utterances. They will be labelled as fluent non-native and non-fluent non-native speakers of English in the following discussion.[1]

By comparing the performance of the two groups, different stages in the development of interlanguage may be observed. Their performances are also compared against native English speakers' English and native Japanese speakers' Japanese

[1]In general the level of proficiency in L2 pronunciation seems to be little correlated with a number of years of exposure to L2. Other variables such as sex, the age at which the learning takes place, motivation and sociolinguistic factors (ethnic or cultural backgrounds) seem to interfere with the time of exposure to L2. Because of this complexity, there is no simple way of placing subjects into different levels of performance. In this experiment, I had to resort to my own impressionistic judgement of the subjects' performance in their fluency or 'native-like-ness' in speaking English. My judgement was supported by an experienced phonetician's judgement as described below.

to see if they would exhibit any sort of 'merged' system, an intermediate stage in interlanguage between L1 and L2. As suggested by Corder (1976), the starting point of interlanguage may be L1. Learners develop their interlanguage through different stages. Earlier stages may be characterized by more L1-like sounds due to a transfer from the L1 system, whereas later stages may be characterized by more L2-like sounds. In the following experiment, the vowel /a/ of Japanese is studied as a possible candidate for tansfer in the production of schwa. It is auditorily most similar to /ə/ among the five vowels /i, e, a, o, u/ of Japanese and the orthography 'a' may also influence the transfer.

In the present study, systematic variability of schwa F2 values as a function of contexts is studied. However, L2 production may also be characterized by non-systematic variability at earlier stages. This early stage may be described as a 'trial and error' stage in which L2 speakers try out different possibilities to approach the quality of the L2 sound.

## 10.2 Experiment 6: General pattern in the production of schwa by Japanese speakers of English

### 10.2.1 Methods

#### Materials

The same nine VCəCV sequences used in Experiment 1 were used in the present experiment. The contextual consonants were /p, t, k/ and the vowels were /ɪ, æ, u/. These 9 sequences were embedded in natural English sentences; e.g., *Please d*ip a pi*n in the solution.* Six non-native speakers of English read these sentences in a sound treated recording studio at the Department of Linguistics, University of Edinburgh. Each sentence was repeated 10 times. In order to compare the Japanese speakers' production of schwa with the native speakers' production, the data from Experiment 1 were used. Part of the data from Experiment 5 was also used as a reference in order to compare the Japanese speakers' production of the non-native schwa and their production of the native (Japanese) vowel /a/. (See page 41 and page 190 for the lists of sentences used in Experiments 1 and 5.)

### Speakers

The data from native speakers of British English, six non-native (Japanese) speakers of English and three native speakers of Japanese were used in this experiment. All of them are male speakers. The three native speakers of British English are AH, MB and DG. Their data are taken from Experiment 1 in Chapter 3. The three native speakers of Japanese are MN, SO, and KO. Their data are taken from Experiment 5 in Chapter 8. The six non-native speakers are KN, MN, MT, TT, HK and HS. KN had an experience of living in the United States from the age of 9 to 13. At the time of the recording, he had been living in Edinburgh for two years. MN had lived one year in the United States from the age of 18 to 19. He is a postgraduate student in phonetics. He had arrived in Edinburgh a month before the recording, and at the time of the recording his English was American accented. MT had lived 3 years in Great Britain and 3 years in Canada from the age of 7 to 12. His English was American accented. He was recruited as a paid volunteer in Tokyo. TT had been living in Edinburgh for three years when the recording was done. HK was a non-graduating student at the University of Edinburgh. At the time of the recording he had been living in Edinburgh for four months. Prior to that he had had no experience of living abroad. HS had been living in Edinburgh for a year at the time of the recording. All the subjects participated in this experiment were either students, staff or visiting fellows at the University of Edinburgh at the time of the recording except MT.

### Analyses and Statistics

For the analysis and statistical methods, see 3.2.1: p 42 for detailed description.

### 10.2.2 Results

Figure 10.1(a) shows a scatterplot of schwa tokens in the 9 VC_CV contexts for the three native speakers of English. The spread is observed mainly in $F_2$ values. Figure 10.1(b) shows a scatterplot of schwa tokens in the same 9 contexts produced by non-native speakers. The variation is observed in two different directions; one in $F_1$ and the other in $F_2$. A careful observation of this figure suggests that these tokens may be broken down into two groups; one group encompassing the production of KN, MN and MT and the other group encompassing the production of TT, HK and HS.

These two groups correspond to the author's impressionistic judgement of fluency among these speakers. KN, MN and MT seem to comprise a group of fluent non-native speakers of English, whereas TT, HK and HS comprise a group of non-fluent non-native speakers. The author's impressionistic judgement was supported by the judgement of an experienced phonetician who is a native speaker of English. He listened to the sentence *You may pick a kitten from the basket* spoken by the nine subjects who participated in this experiment. He was not told that there were sentences produced by native speakers among the stimuli. He was asked to rate their fluency in the scale of 1 to 10. The speakers TT, HK and HS were given the scores of 2, 3 and 4 respectively, while the speaker MT was given the score of 5 points and KN and MN scored 6. The native speakers' production scored 7 and 9. DG scored the highest, probably because his pronunciation was closest to the RP system.

The difference in the production of fluent and non-fluent speakers is observed in the direction of the spread of their tokens. Fluent speakers show a spread of schwa in $F_2$ values, whereas non-fluent speakers show a spread in $F_1$ values. Figure 10.2 shows this trend. The difference in standard deviations illustrates the different directions in which fluent and non-fluent native speakers' schwas spread. The standard deviations for the $F_1$ values of schwa were 73 Hz and 196 Hz for fluent and non-fluent speakers of English respectively, showing greater variation for non-fluent speakers. On the other hand, the standard deviations for $F_2$ were 290 Hz and 154 Hz for fluent and non-fluent speakers, suggesting that while fluent speakers showed the $F_2$ variation characteristic of schwa, non-fluent

```
          (a)  Native speakers' schwas
          .....+.....+.....+.....+.....+.....+..
          -                                    -
          -                                    -
    2500  +                                    +
          -                                    -
          -                                    -
          -        B                           -
f         -     A C B                          -   A = AH
2   2000  +      A**BB                         +   B = NB
          -     AA***B                         -   C = DG
          -      A**BB                         -
          -      ****                          -
          -     A***B B                        -
    1500  +     A**** *                        +
          -      *A***                         -
          -  A   C **A*BB                      -
          -       CCB*                         -
          -      CC*CC                         -
    1000  -       CC                           -
          -                                    -
          -                                    -
          -                                    -
          .....+.....+.....+.....+.....+.....+..
              200         600         1000
                    400         800         1200

          (b)  Non-native speakers' schwas
          .....+.....+.....+.....+.....+.....+..
          -                                    -
          -                                    -
    2500  +       F                            +   D = KN
          -         F                          -   E = MN
          -      F FF                          -   F = MT
          -        FF                          -   G = TT
f         -       D**FF                        -   H = HK
2   2000  +          FF                        +   I = HS
          -         FF F                       -
          -      D HF HF                       -
          -      ********HH          HH        -
          -      I*********        HH  H    G  -
    1500  -    G **********      *G    G    G  +
          -    II*********EEI I IH    *        -
          -    GI**********E** H*IH*HHH        -
          -      ******D** * I                 -
          -    D  DFD*                         -
    1000  +       *                            +
          -                                    -
          -        F                           -
          -                                    -
          .....+.....+.....+.....+.....+.....+..
              200         600         1000
                    400         800         1200

                           f1
```

**Figure 10.1.** The scatterplot of schwas produced by (a) native and (b) non-native speakers of English. Symbol * indicates an overlap between tokens belonging to different speakers.

speakers did not.

When schwa tokens produced by native and fluent non-native speakers are plotted on the same graph, the similarity between their pattern of schwa production becomes clear (Figure 10.3). The spread in $F_2$ values as a function of contexts is similar between the native and non-native speakers' production of schwa. The non-native speakers' schwas are generally higher in $F_1$ values than the native speakers' schwas. The mean F1 value of schwa produced by fluent non-native speakers of English was 389 Hz, while the mean $F_1$ value produced by native speakers of English was 292 Hz. There was a significant difference in $F_1$ values of schwa ($F(1,492) = 310.01$, $p < 0.0001$). The standard deviations for both $F_1$ and $F_2$ values were greater for fluent non-native than native speakers: 46 Hz for native and 73 Hz for non-native speakers in $F_1$, and 250 Hz for native and 290 Hz for non-native speakers of English in $F_2$. Non-native speakers' schwas seem to be more variable in $F_1$. For $F_2$, the extreme variability observed for MT seems to be affecting the group value.

Table 10.1 shows the matrices of the mean $F_1$ and $F_2$ values of schwa in the 3 consonant $\times$ 3 vowel contexts for each group of speakers, native, fluent non-native and non-fluent non-native speakers of English and native speakers of Japanese. The data for the Japanese vowel /a/ are taken from Experiment 5 in Chapter 8.

When the mean $F_1$ values for each group are compared, they increase in the following order: native English < fluent non-native English < non-fluent non-native English < native Japanese. The values for the non-native speakers are intermediate between the mean $F_1$ values of the Japanese vowel /a/ and the native speakers' schwa, suggesting some sort of 'merged' system. Further, the fluent non-native speakers' value was closer to the native speakers' value of schwa. Significant difference was observed between the $F_1$ values of the four groups of speakers ($F(3,934) = 391.04$, $p< 0.0001$). The results of the post hoc scheffe tests show that the differences in the mean $F_1$ values between any two of these groups are significant by $p < 0.01$.

A similar pattern is observed for the mean $F_2$ values. They increase in the following order: native Japanese < non-fluent non-native English < fluent non-native English < native English. Significant difference was observed between the

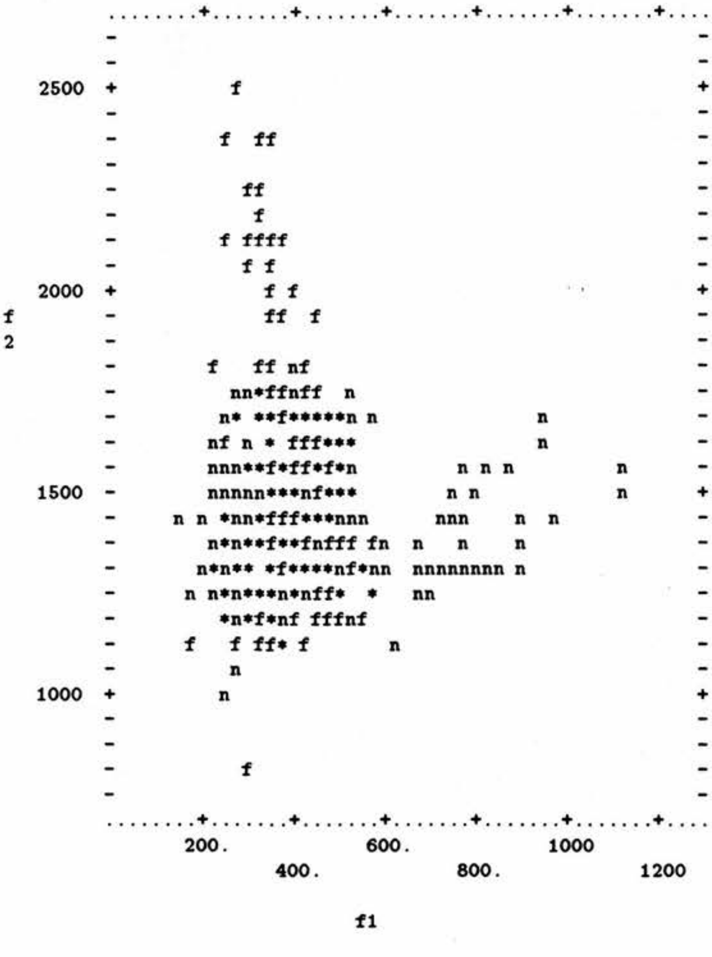



**Figure 10.2.** The scatterplot of schwa produced by fluent and non-fluent non-native speakers of English. Symbols f = fluent, n = non-fluent, * = an overlap between tokens belonging to different groups.

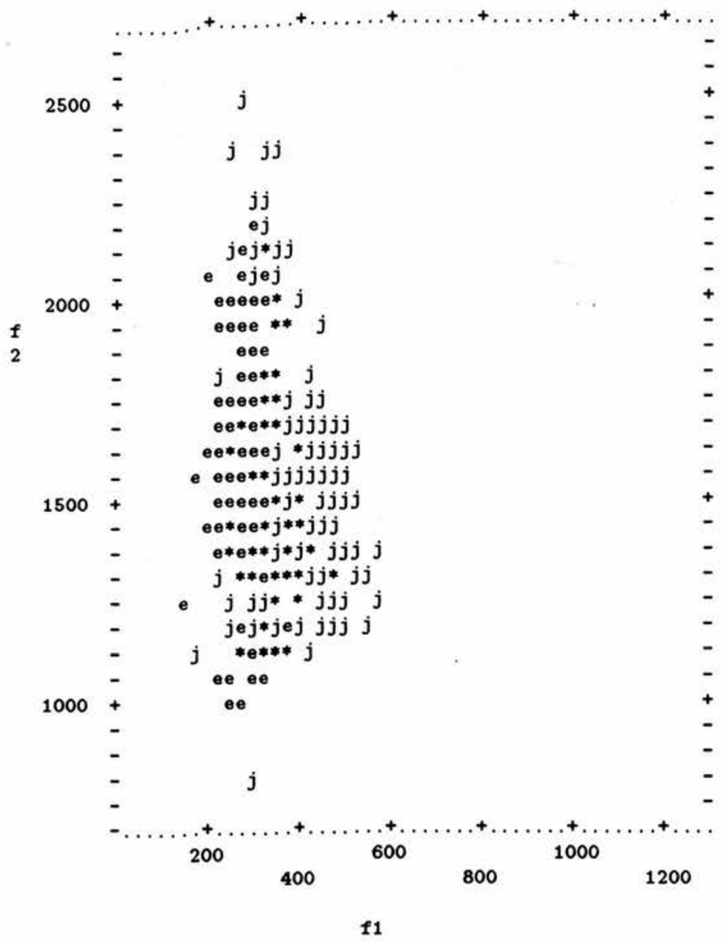



**Figure 10.3.** The scatterplot of schwa produced by native speakers and fluent non-native speakers of English. Symbols e = native, j = fluent non-native, * = an overlap between tokens belonging to different groups.

| $F_1$ | | | | | $F_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Native English | p | t | k | | | p | t | k | |
| ɪ | 301 | 276 | 284 | 287 | ɪ | 1391 | 1659 | 1945 | 1672 |
| æ | 321 | 283 | 293 | 296 | æ | 1266 | 1567 | 1850 | 1561 |
| u | 312 | 275 | 286 | 290 | u | 1263 | 1640 | 1562 | 1491 |
| | 310 | 278 | 288 | 291 | | 1306 | 1621 | 1786 | 1573 |
| Fluent non-native | p | t | k | | | p | t | k | |
| ɪ | 403 | 391 | 361 | 384 | ɪ | 1366 | 1598 | 1858 | 1607 |
| æ | 390 | 448 | 353 | 397 | æ | 1306 | 1484 | 1805 | 1532 |
| u | 406 | 379 | 376 | 386 | u | 1211 | 1474 | 1472 | 1386 |
| | 399 | 406 | 362 | 389 | | 1294 | 1512 | 1716 | 1508 |
| Non-fluent non-native | p | t | k | | | p | t | k | |
| ɪ | 323 | 328 | 403 | 354 | ɪ | 1319 | 1570 | 1501 | 1463 |
| æ | 350 | 390 | 341 | 362 | æ | 1269 | 1536 | 1380 | 1395 |
| u | 656 | 395 | 506 | 515 | u | 1297 | 1452 | 1398 | 1382 |
| | 458 | 371 | 414 | 411 | | 1295 | 1519 | 1427 | 1413 |
| Native Japanese | p | t | k | | | p | t | k | |
| i | 682 | 620 | 637 | 645 | i | 1319 | 1567 | 1469 | 1453 |
| a | 667 | 552 | 577 | 596 | a | 1159 | 1484 | 1380 | 1349 |
| u | 610 | 603 | 588 | 600 | u | 1201 | 1453 | 1311 | 1323 |
| | 655 | 591 | 601 | 614 | | 1229 | 1502 | 1386 | 1375 |

**Table 10.1.** The $F_1$ and $F_2$ values at the midpoint of schwa in the 3 consonant /p, t, k/ × 3 vowel /ɪ, æ, u/ (/i, a, u/) contexts for each group of speakers. The rightmost column of each matrix shows the mean formant frequencies for the vocalic contexts. The bottom row shows the mean formant frequencies for the consonantal contexts. The rightmost value on the bottom row is the grand mean for each matrix.

$F_2$ values of the four groups of speakers ($F(3,934) = 91.57$, $p < 0.0001$). However, the results of the post hoc scheffe test have shown no significant difference between the non-fluent non-native speakers' schwa and the Japanese vowel /a/. Also, the difference in the $F_2$ values of schwa produced by native speakers and fluent non-native speakers was significant only by $p < 0.05$. The difference was marginal.

When we study the trend observed for $F_2$ values in these matrices, it becomes clear that the fluent non-native speakers' schwas pattern very much like those of native speakers. On the other hand, non-fluent non-native speakers' schwas pattern more like the Japanese vowel /a/.

First of all, the range of variation from the /up_pu/ to /ɪk_kɪ/ contexts is similar between native and fluent non-native speakers. They range 1263 Hz to 1945 Hz for native speakers and 1211 Hz to 1858 Hz for fluent non-native speakers in mean value. The range for non-fluent non-native speakers of English is more like the range observed for the Japanese vowel /a/.[2] They are 1297 Hz (/up_pu/) to 1501 Hz (/ɪk_kɪ/) for the schwa production, and 1229 Hz (/up_pu/) to 1491 Hz (/ik_ki/) for the production of the Japanese vowel /a/ in mean value.

Secondly, for native and fluent non-native speakers, $F_2$ values increase according to consonantal contexts, $p < t < k$. On the other hand, the $F_2$ values of the non-fluent non-native speakers' schwas and the $F_2$ values of the Japanese vowel /a/ increase according to consonantal contexts, $p < k < t$.

The data suggest that the pattern observed for non-fluent non-native speakers is very much that of their native language vowel /a/. In other words, there is clearly a transfer from L1 to the L2 production of schwa. They have not acquired the important characteristic of schwa which is its targetlessness in $F_2$. The range of the $F_2$ values of schwa produced by non-fluent speakers is 776 Hz from the minimum (1007 Hz) to the maximum (1783 Hz) value. They do not show the spread in $F_2$ values as a function of contexts which is an important feature of the English schwa. While the spread of schwa tokens in $F_2$ value is very much like that of the Japanese vowel /a/, the spread in $F_1$ value is far greater for their schwa than for the vowel /a/ of Japanese.

[2]Subjects were not matched for the two sets of data. The English schwa production and the Japanese vowel /a/ were produced by different groups of native speakers of Japanese. MN is the only subject who produced both sets. His data will be discussed later.

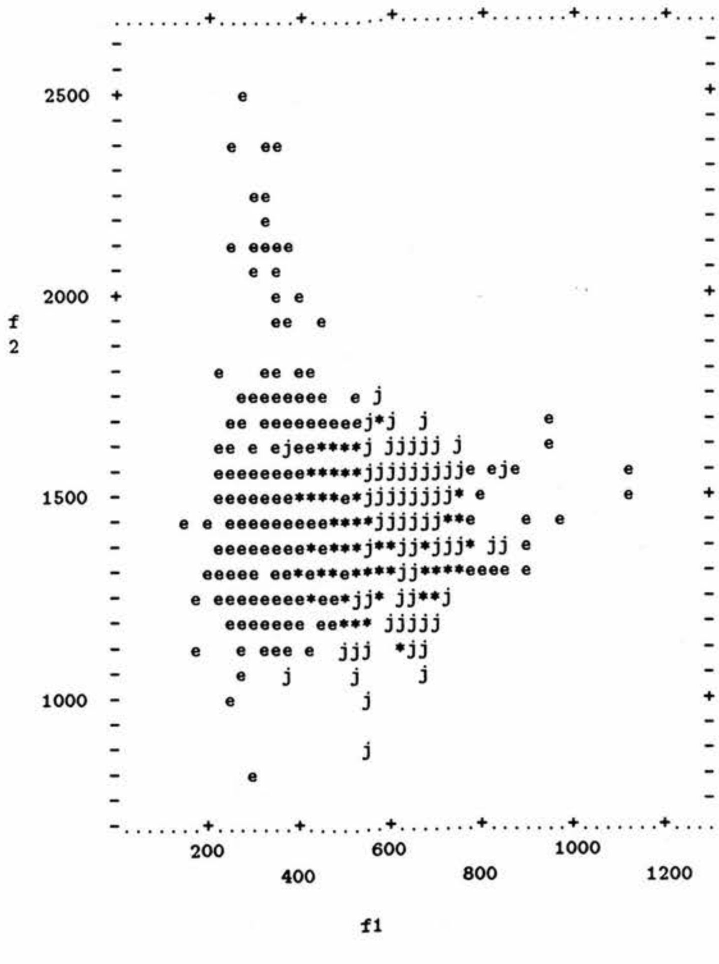



**Figure 10.4.** The scatterplot of schwa produced by non-native speakers of English and the Japanese vowel /a/. Symbols e = English schwa, j = Japanese vowel /a/, * = an overlap between tokens belonging to different groups.

| | | AH | MB | DG |
|---|---|---|---|---|
| | | $R^2$ | $R^2$ | $R^2$ |
| $F_2$ | C | 0.5262 | 0.7415 | 0.6386 |
| | $V_1$ | 0.2643 | 0.1487 | 0.2019 |
| | $V_2$ | 0.1908 | 0.0415 | 0.0255 |
| | $V_1 + V_2$ | 0.3265 | 0.1598 | 0.0255 |
| | $C + V_1$ | 0.6565 | 0.7557 | 0.7064 |
| | $C + V_2$ | 0.8641 | 0.8928 | 0.8671 |
| | $C + V_1 + V_2$ | 0.8742 | 0.8932 | 0.8712 |

**Table 10.2.** The results of the multiple regression analyses for the $F_2$ variability at the midpoint of the vowel /ə/. The independent variables are consonantal context (C) and the vocalic context (V). For consonantal contexts, arbitrary scores of 1 to 3 were given for /p, t, k/, and for vocalic contexts, the $F_2$ values at the midpoint of the preceding ($V_1$) and the following ($V_2$) vowel were used.

When the Japanese vowel /a/ tokens in the same 9 VC_CV contexts are overlaid on top of the scatterplot of /ə/ produced by non-native speakers, it becomes apparent that the Japanese tokens and English tokens produced by Japanese speakers do not overlap completely. Their schwa is very unstable and variable in $F_1$. The spread in $F_1$ value observed for non-fluent non-native speakers seems to be a result of non-systematic variability rather than a transfer. The extensive variability observed in $F_1$ for the production of schwa is not an inherent characteristic of the Japanese vowel /a/. This extensive variability seems to arise as a result of uncertainty in production.

On the other hand, fluent non-native speakers of English seem to have acquired the important characteristic of schwa, i.e., its targetlessness in $F_2$. Their schwas vary systematically as a function of context in $F_2$. The magnitude of variation is in fact greater than that of the native speakers. The range of the $F_2$ values of schwa is 1169 Hz from the minimum (1002 Hz) to the maximum (2171 Hz) value for the native speakers of English, whereas the range is 1641 Hz from the minimum (844 Hz) to the maximum (2485 Hz) value for the fluent non-native speakers of English. This is due to the greater degree of variability observed for MT.

| | | KN | MN | MT | TT | HK | HS |
|---|---|---|---|---|---|---|---|
| | | $R^2$ | $R^2$ | $R^2$ | $R^2$ | $R^2$ | $R^2$ |
| $F_2$ | C | 0.4871 | 0.5839 | 0.5449 | 0.1054 | 0.2398 | 0.0959 |
| | $V_1$ | 0.0493 | 0.2347 | 0.0262 | 0.3214 | 0.0898 | 0.0447 |
| | $V_2$ | 0.0239 | 0.1920 | 0.4576 | 0.2391 | 0.0082 | 0.0188 |
| | $C + V_1$ | 0.5717 | 0.7006 | 0.5484 | 0.3919 | 0.2810 | 0.1248 |
| | $C + V_2$ | 0.6175 | 0.7755 | 0.6713 | 0.3895 | 0.2507 | 0.1264 |
| | $C + V_1 + V_2$ | 0.6182 | 0.7762 | 0.6817 | 0.4095 | 0.3460 | 0.1269 |

**Table 10.3.** The results of the multiple regression analyses for the $F_1$ and $F_2$ variability at the midpoint of the vowel /ə/. The independent variables are consonantal context (C) and the vocalic context (V). For consonantal contexts, arbitrary scores of 1 to 3 were given for /p, t, k/, and for vocalic contexts, the $F_2$ values at the midpoint of the preceding ($V_1$) and the following ($V_2$) vowel were used.

The results of multiple regression analyses show the same trend. When the $r^2$ values for the combination of the variables $C + V_1 + V_2$ are compared for the three groups of subjects, they are higher for natives speakers: 0.8742 (AH), 0.8932 (MB) and 0.8712 (DG) (Tables 10.2 and 10.3). The $r^2$ value is the proportion of variability accounted for by the variables included in the analysis among the total variability observed in the data. Thus, the combination of the above three variables accounted 87.42 % of the total variability of the $F_2$ values of schwa produced by AH, and so on. The $r^2$ values are smaller for non-fluent non-native speakers: 0.4095 (TT), 0.3460 (HK) and 0.1269 (HS). The $r^2$ values for fluent non-native speakers are intermediate in value between the values observed for the other two groups: 0.6182 (KN), 0.7762 (MN) and 0.6817 (MT). This seems to be an interesting case of 'interlanguage continuum' (Ellis 1994) where speakers in a relatively advanced stage of L2 acquisition manifest values intermediate in degree between those observed for native speakers of L2 and non-native speakers who are in a less advanced stage.

MN was the only subject who produced both schwa and the Japanese vowel /a/ in the 9 VC_CV contexts. The mean $F_1$ and $F_2$ values for his Japanese vowel /a/ were 640 Hz and 1334 Hz, while the mean $F_1$ and $F_2$ values for his English schwa were 454 Hz and 1496 Hz. Figure 10.5 shows his data. His

| | | MN | |
|---|---|---|---|
| | | Japanese | English |
| $F_1$ | C | 0.2315 | 0.3214 |
| | V | 0.0194 | 0.0000 |
| | C + V | 0.2400 | 0.3237 |
| $F_2$ | C | 0.0244 | 0.5568 |
| | V | 0.1536 | 0.0511 |
| | C + V | 0.1920 | 0.6368 |

**Table 10.4.** The results of the multiple regression analyses for the $F_1$ and $F_2$ variability at the midpoint of the Japanese vowel /a/ and the English /ə/produced by a Japanese speaker of English (MN). The independent variables are consonantal context (C) and the vocalic context (V). For consonantal contexts, arbitrary scores of 1 to 3 were given for /p, t, k/, and for vocalic contexts, scores of 1 to 3 were given for /i, a, u/ (/ɪ, æ, u/).

schwa tokens are clearly different from the Japanese vowel /a/ tokens, and they show the characteristic spread of the $F_2$ values of schwa. His Japanese vowel /a/ is also quite variable in both $F_1$ and $F_2$. However, the distribution of the two vowel groups is clearly different. The results of the regression analyses also showed difference between his English /ə/ and his Japanese /a/ (Table 10.4).[3] The difference is most obvious for $F_2$; the $r^2$ value of 0.1920 vs. 0.6368 for Japanese and English respectively. His schwa exhibits greater effects of context than does his Japanese vowel /a/. Also, his English schwa showed greater effects of consonantal contexts than vocalic contexts, while his Japanese /a/ showed greater effects of vocalic than consonantal contexts. On the other hand, the $r^2$ values for $F_1$ were similar between his Japanese and English tokens.

Another possible source for the difference observed in the production of schwa by the three groups of speakers may be difference in the duration of schwa. As contextual assimilation is correlated with duration (Lindblom 1963), more systematic variability as a function of contexts may be observed when the segment

[3]As many of the contextual vowels in the VCaCV utterances of Japanese were devoiced and the $F_2$ values were not obtained, the arbitrary scores of 1 to 3 had to be used for both the Japanese vowels /i, a, u/ and the English vowels /ɪ, æ, u/ rather than actual $F_2$ values as independent variables in the regression analyses.

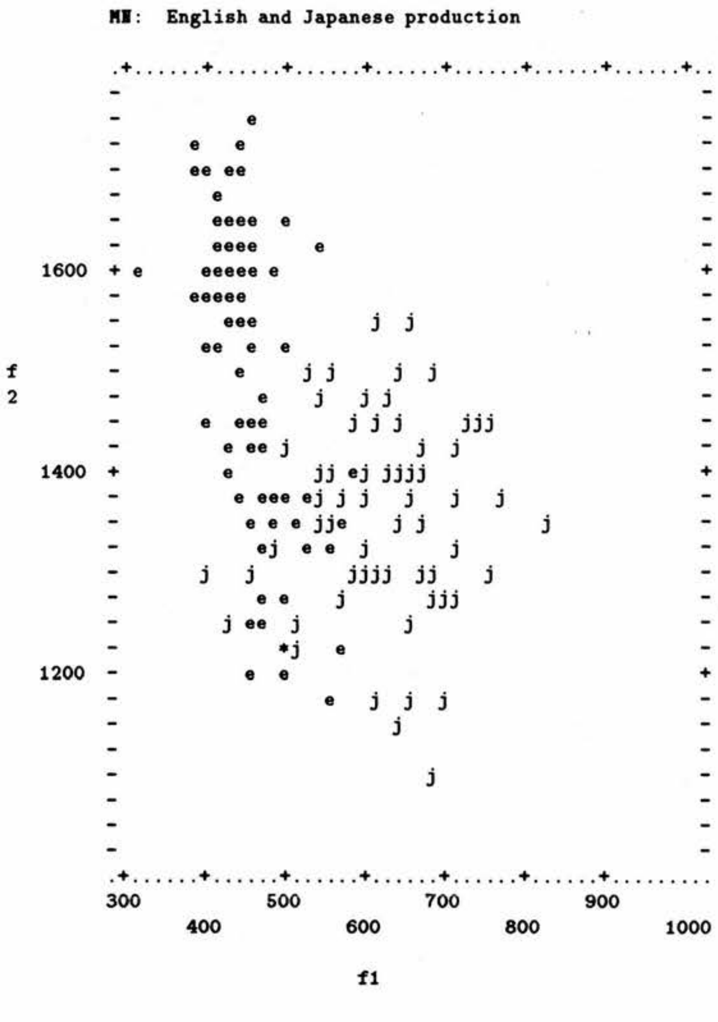



**Figure 10.5.** The scatterplot of the English schwa and the Japanese vowel /a/ produced by MN. Symbols e = English schwa, j = Japanese vowel /a/, * = an overlap between tokens belonging to different groups.

| $F_1$ | | | | | $F_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Japanese /a/ | p | t | k | | | p | t | k | |
| i | 778 | 663 | 720 | 718 | i | 1331 | 1503 | 1404 | 1415 |
| a | 785 | 674 | 604 | 693 | a | 1209 | 1423 | 1319 | 1339 |
| u | 741 | 617 | 588 | 635 | u | 1226 | 1421 | 1394 | 1353 |
| | 773 | 652 | 644 | 686 | | 1265 | 1450 | 1372 | 1371 |
| English /ə/ | p | t | k | | | p | t | k | |
| ɪ | 498 | 446 | 439 | 461 | ɪ | 1388 | 1571 | 1684 | 1548 |
| æ | 508 | 490 | 442 | 480 | æ | 1368 | 1550 | 1643 | 1520 |
| u | 540 | 429 | 431 | 455 | u | 1279 | 1598 | 1500 | 1459 |
| | 511 | 455 | 437 | 466 | | 1345 | 1573 | 1609 | 1509 |

**Table 10.5.** The $F_1$ and $F_2$ values at the midpoint of schwa in the 3 consonant /p, t, k/ $\times$ 3 vowel /ɪ, æ, u/ (/i, a, u/) contexts for the /ə/ and the Japanese /a/ produced by MN. The rightmost column of each matrix shows the mean formant frequencies for the vocalic contexts. The bottom row shows the mean formant frequencies for the consonantal contexts. The rightmost value on the bottom row is the grand mean for each matrix.

is shorter. In fact, native speakers' schwas were the shortest among the three groups (35 ms), while non-fluent non-native speakers' schwas were relatively long (70 ms). Fluent non-native speakers' schwas were again intermediate in value between the other two groups (42 ms). The difference between these three groups was significant ($F(2,800) = 471.57$, $p < 0.0001$). The result of the post hoc scheffe test shows that differences between every combination of these means are significant by $p < 0.01$.

On the other hand, the duration of schwa and the Japanese vowel /a/ produced by MN did not show any significant difference ($F(1,177) = 2.88$, $p = 0.0912$) with the mean duration of 42.9 ms (SD = 12.8 ms) for schwa and the mean duration of 46.3 ms (SD = 13.6 ms) for the vowel /a/. However, the patterns observed for the distribution of his /ə/ and /a/ are different as shown in Figure 10.5. This supports the view that though duration may affect the extent of contextual assimilation, the underspecification of schwa may be phonologically determined. MN seems to have acquired this phonological feature of schwa in the course of learning English. Table 10.5 shows the shift from L1 to a native-like pattern in the production of schwa for MN. His schwas do not vary as extensively as native speakers' schwas in $F_2$. His spread in $F_2$ values as a function of context is also smaller in magnitude compared to the other two fluent non-native speakers KN and MT (see Appendix C). However, the pattern observed for his schwa is clearly closer to the native speakers' pattern of schwa than to the pattern of the Japanese vowel /a/. (See also Table 10.1.)

### 10.2.3 Summary of results

1. Schwas produced by non-native speakers of English showed significant difference between fluent and non-fluent speakers of English. These two groups may represent different stages in the developmental process of interlanguage.

2. The production of schwa by non-fluent speakers of English may be characterized by a transfer from their native language vowel /a/ in $F_2$ dimension, and non-systematic and large variability in $F_1$ dimension, which is not observed in the production of the Japanese vowel /a/ but is unique to their schwa production.

3. The production of schwa by fluent non-native speakers of English may be characterized as being very similar to the native speakers' production of schwa. Their schwas vary systematically and largely as a function of contexts. They seem to have acquired the important feature of schwa, its targetlessness in $F_2$. Their schwas are however higher and more variable in $F_1$ values compared to the native speakers' schwas.

4. The mean $F_1$ and $F_2$ values of schwa produced by non-native speakers of English were intermediate in value between those observed for native speakers' schwa and those observed for the Japanese vowel /a/ of the L1, suggesting a 'merged' system. Further, the mean values observed for fluent non-native spakers of English were closer to the native speakers' values, suggesting different stages in the interlanguage of schwa.

5. The results of the multiple regression analyses suggest that the systematicity of schwa variation as a function of contexts is most pronounced for the native speakers' production of schwa while it is least pronounced for non-fluent English speakers' schwas. There was an intermediate degree of systematicity observed for fluent non-native speakers' schwas.

### 10.2.4 Discussion

Non-fluent speakers of L2 did not show extensive variation in $F_2$ as a function of contexts. That no significant difference in $F_2$ value was observed between the non-native /ə/ and the native /a/ suggests that this stage may be characterized by a transfer from the L1 vowel system. However, large variability in $F_1$ which is unique to the production of the non-native /ə/ seems to suggest that L2 speakers are aware of the difference between the non-native /ə/ and the native /a/, trying out random productions to approach the quality of /ə/.

L2 speakers at an advanced stage seem to have managed to acquire the important characteristic of schwa, its phonetic underspecification in $F_2$. It is rather unlikely that the speakers KN, MN and MT had access to the explicit knowledge of phonetic underspecification of schwa.[4] In other words, they would not have been told that schwa is targetless. Their knowledge of phonetic underspecification of schwa must be implicit in nature.

Then a question arises: where does this knowledge come from? One possibility is that they perceive the coarticulatory pattern in the L2 stimuli and eventually develop the pattern in their interlanguage. While L2 learners may not have access to the 'explicit' knowledge of the phonetic underspecification of schwa, they may have some explicit knowledge of the rhythmic alternation of strong and weak syllables in English. This feature of the English rhythm may also be relatively easy to perceive. By producing schwa as a weak and short vowel, L2 learners may succeed in producing a more contextually variable and therefore targetless schwa. Production of targetless schwa may consequently help them to make more prominent contrast between reduced and full vowels in English, resulting in a more

[4]MN was a postgraduate student in phonetics at the University of Edinburgh, but he was not aware of the purpose of my study.

English-like rhythm. This, in turn, may encourage them to make a less and less specified schwa. With the help of some explicit knowledge or external stimuli, L2 learners may develop the phonetic features of their interlanguage towards a more L2-like system.

The different realizations of /ə/ and the Japanese vowel /a/ by MN suggest that L2 learners may eventually develop the implicit knowledge of the phonetic underspecification of schwa. Somewhere in the developmental process of interlanguage, there seems to be a shift where what is phonetic in nature becomes part of the grammar of the interlanguage. This process may reflect how a language picks up or emphasizes certain phonetic features in the grammar of the language. For example, English seems to be using contextual assimilation as part of its phonology in order to make a contrast between stressed and unstressed vowels. Also, vowel devoicing which seems to be a universal vowel weakening process is emphasized in Japanese. Hirayama (1985:p44) suggests in "Zen Nihon no hatuon to akusento" (the Pronouncing Dictionary of Japanese) that a moderate amount of devoicing is a desirable feature in Standard Japanese. Speakers also seem to be able to control the extent of coarticulation with training. Habis (forthcoming) compares the spread of certain features, such as [+pharyngeal] onto vowels between trained expert Koran reciters and ordinary speakers of Arabic. He observes a more abrupt transition from an emphatic consonant to the following vowel for experts whereas the transition is more gradual for ordinary speakers.

Without a longitudinal study we may not conclude that the two different patterns observed for the fluent and non-fluent speakers of L2 in the present study represent different developmental stages in the interlanguage. However, two clearly different patterns, one which is better characterized as a transfer from the L1 system and the one which is more like the L2 system, were observed.

The two subjects KN and MT were exposed to English at a relatively early stage in their life. They spent 6 years during their primary education period in an English speaking environment. This may have affected their production of schwa. However, MN who had not lived in an English speaking environment till the age of 18 did a very good job in the production of schwa. This is a very encouraging fact from a pedagogical point of view. The transparency of schwa may be an important feature of the rhythmic organization of English as discussed in Part I. Therefore, its correct acquisition seems to be essential to the acquisition of the rhythm of English. The results of the present study are encouraging in that L2 speakers seem to eventually develop the right coarticulatory pattern of schwa. By introducing this feature of schwa as part of conscious learning, the acquisition of schwa, and consequently the acquisition of the rhythm of English, may be assisted and enhanced.

## 10.3 Conclusion

In the present study, the production of schwa by two groups of non-native speakers was observed. Two different coarticulatory patterns were observed among the production of schwa by non-native speakers of L2. While the non-fluent non-native speakers did not show the characteristic variability of schwa in $F_2$, fluent non-native speakers' schwa varied systematically as a function of contexts. Their schwa patterned very much like native speakers' schwa. They seem to have learned the phonetic underspecification of schwa in $F_2$.

# Chapter 11

# Interlanguage of Schwa 2

## 11.1 Introduction

In the previous chapter it was shown that while non-fluent speakers exhibited a coarticulatory pattern which seems to be a transfer from the L1 vowel /a/ in $F_2$ and non-systematic variability in $F_1$, fluent speakers showed a more L2-like schwa. Thus, non-native speakers of English in an advanced stage seem to have acquired the phonetic underspecification of schwa. They have shown a shift from the L1 to L2 coarticulatory pattern.

In the present experiment, I will further explore the nature of the interlanguage of schwa by comparing the V-to-V effects on the English /ə/ and the full vowel /æ/ produced by Japanese speakers of English. In Experiment 2, the V-to-V effects on the /ə/ and /æ/ produced by native speakers of English were compared. In general native speakers of English showed a greater extent of V-to-V effects on /ə/ than on /æ/ both in magnitude and in temporal extent. It was suggested that this contrast in the transparency of a reduced and a full vowel may be an important feature of stress-timing.

In Experiment 6, the non-native speakers of English showed a merged system in the production of /ə/. That is, both the $F_1$ and $F_2$ values of schwa as well as the amount of context dependent variability were intermediate in value between the native speakers' schwa and the Japanese vowel /a/. The effects of the combination of consonant and vowel contexts on the $F_2$ values of schwa became increasingly greater through non-fluent non-native < fluent non-native < native.

In comparing the non-native speakers' coarticulatory patterns for the English vowels /ə/ and /æ/, the merged system hypothesis (Flege & Hillenbrand 1984) would predict that the contrast of targetedness in $F_2$ may be less pronounced for non-native speakers than for native speakers of English. In other words, Japanese speakers of English may manifest a contrast between the full and reduced vowels /æ/ and /ə/ of English as a function of contexts. However, the difference in the magnitude of context dependent variability between /æ/ and /ə/ would be smaller for non-native than native speakers of English.

- Hypothesis: In shifting from the L1 system to the L2 system, speakers of L2 will show a merged system, that is, a system which is intermediate in value between the two systems. In the present case, Japanese speakers of English will manifest different degrees of context dependent variability between the English full vowel /æ/ and the reduced vowel /ə/. However, the difference in the magnitude of variability between the two types of vowels will be smaller than the difference observed for native speakers' production.

## 11.2 Experiment 7

### 11.2.1 Methods

The English Vb_bV utterances with /ə/ and /æ/ as the middle vowel were produced by five male Japanese subjects. The contextual vowels were /ɪ/ and /æ/ (/ə/) (see page 76 for the description of the materials). These data were compared with the same English Vb_bV utterances produced by eight native British English speakers (Experiment 2) and the Japanese Vb_bV utterances (the vowel /a/ as the middle vowel and the vowels /i/ and /a/ as the contextual vowels) produced by the same five Japanese subjects above (Experiment 4b: see page 171 for the description of the Japanese materials). There were both symmetric and asymmetric contexts resulting in four different sentences for each of the vowels /ə/, /æ/ and the Japanese vowel /a/. Each sentence was repeated five times by each subject. There were unequal number of subjects between the native English speakers' data and the Japanese speakers' data. In order to make the number of observations more or less equal, the middle three tokens from the five repetitions from each native English speaker's data were taken. Thus, there were at most 25 observations (5 speakers $\times$ 5 repetitions) per each sentence type for the data produced by the Japanese subjects and 24 observations (8 speakers $\times$ 3 repetitions) for the native English speakers' data.

**Speakers**

The five Japanese subjects who produced the English Vb_bV utterances were KN, MN, SO, KO and SK. KN and MN participated in Experiment 6 as well. They were both scored 6 in the overall fluency of their English production and judged as fluent non-native speakers. SO had been living in Edinburgh for over four years at the time of the recording. KO had lived in Wales for one year. SK had been living in Edinburgh for one year at the time of the recording. Apart from KN, the other four subjects were either language teachers and/or graduate students in linguistics and had a certain amount of interest in languages. I considered them to be relatively fluent speakers of English.[1]

### 11.2.2 Results and discussion

Three-way ANOVAs were performed for the English data produced by the Japanese speakers. The dependent variables were $F_2$ values of the English vowels /ə/ and /æ/ at the onset, midpoint and offset of the segment. The independent variables were the preceding vowel, following vowel and speaker. Table 11.1 summarizes the results. No significant effects of the preceding vowel were observed at any point of the segment for either /ə/ or /æ/, whereas significant effects of the following vowel were observed right through /ə/ and /æ/. Significant interactions were observed between the following vowel and speaker throughout the segment for /ə/ and at the midpoint and offset of the segment for /æ/. In these cases the five speakers showed different degrees of the following vocalic effects. KN had higher mean $F_2$ values for schwa in the context of /æ/ than in that of /ɪ/ at all

[1]Possibly due to some sociolinguistic factors, it was difficult to find male Japanese subjects whose pronunciation of English was reasonably good. This resulted in the small number of subjects in the present study.

| | ə | | | æ | | |
|---|---|---|---|---|---|---|
| | onset | midpoint | offset | onset | midpoint | offset |
| pre_v (p) | 0.0861 | 0.7337 | 0.8525 | 0.1378 | 0.9879 | 0.1246 |
| fol_v (f) | 0.0039 | 0.0001 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| speaker (s) | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0000 | 0.0016 |
| p × f | 0.7915 | 0.5259 | 0.2296 | 0.7839 | 0.6434 | 0.3146 |
| p × s | 0.1057 | 0.4416 | 0.7576 | 0.1401 | 0.6077 | 0.5138 |
| f × s | 0.0001 | 0.0000 | 0.0000 | 0.1430 | 0.0024 | 0.0039 |
| p × f × s | 0.7637 | 0.2944 | 0.0400 | 0.0978 | 0.1478 | 0.8203 |

**Table 11.1.** The level of significance of the results of ANOVAs. The table shows the main effects of the preceding vowel, following vowel and speaker, and their interaction on the second formant values of schwa and the vowel /æ/ across five speakers.

the three points of the segment. The interaction between the preceding vowel, following vowel and speaker was also significant at the onset of /æ/.[2]

Table 11.2 is a summary of the results of the ANOVAs for the effects of the preceding and following vowel for the English /ə/ and /æ/ produced by native British English speakers and Japanese speakers of English, and for the Japanese vowel /a/ produced by native Japanese speakers. The '+' sign indicates that the effects were significant while the '−' sign indicates otherwise. This table is a summary of Tables 4.1 and 7.3 and 11.1.

Post hoc scheffe tests were also performed for the interaction of the 5 vowel types (native speakers' /ə/ and /æ/, non-native speakers' /ə/ and /æ/ and the Japanese /a/), the 3 points of measurement and the 2 preceding vowels (/ɪ/ and /æ/ or /i/ and /a/). Significant difference was observed only at the onset of the native speakers' schwa as a function of the preceding vowel. For the interaction of the 5 vowel types, the 3 points of measurement and the 2 following vowels

[2]Appendix A shows the mean $F_2$ values as a function of the preceding and following vowel at the onset, midpoint and offset of the English vowels /ə/ and /æ/ produced by each native and non-native speaker of English and the Japanese vowel /a/ produced by each native Japanese speaker. Appendix B shows the mean $F_2$ values in the following four contexts, /ɪb_bɪ/ (/ib_bi/), /ɪb_bæ/ (/ɪb_bə/ or /ib_ba/), /æb_bɪ/ (/ab_bi/) and /æb_bæ/ (/æb_bə/ or /ab_ba/) at the three points of measurement for the English /ə/ and /æ/ produced by native and non-native speakers and the Japanese vowel /a/ produced by native Japanese speakers.

| | | CARRYOVER | | | ANTICIPATORY | | |
|---|---|---|---|---|---|---|---|
| | | onset | midpoint | offset | onset | midpoint | offset |
| English | ə | + | + | + | + | + | + |
| | æ | + | + | − | − | − | + |
| Japanese | ə | − | − | − | + | + | + |
| | æ | − | − | − | + | + | + |
| | a | + | + | + | + | + | + |

**Table 11.2.** The summary of the results of ANOVAs for the vowels /ə/ and /æ/ produced by native British English speakers and Japanese speakers, and for the Japanese vowel /a/ produced by Japanese speakers.

(/ɪ/ or /i/ and /æ/, /ə/ or /a/), significant differences were observed only at the offset of the Japanese vowel /a/ and the English /æ/ produced by the Japanese subjects. The results of the post hoc tests seem to suggest strong carryover effects for the English speakers' schwa and strong anticipatory effects for the Japanese speakers' production of vowels.

A number of interesting observations may be made from Table 11.2. First of all, the reduced vowel /ə/ is more transparent than the full vowel /æ/ when they are produced by native speakers. Secondly, the Japanese vowel /a/ is as transparent as the native speakers' schwa. Thirdly, the Japanese subjects did not show any difference in sensitivity to context between the production of /ə/ and /æ/. Lastly, when the Japanese subjects produced the non-native vowels /ə/ and /æ/, their preference for anticipatory effects seem to be much more pronounced. Somehow their coarticulatory pattern is distorted and different from either the L1 or the L2 pattern. There seems to be a strong over-projection of the preferred coarticulatory pattern of the L1 onto the interlanguage system.

Figures 11.1 and 11.2 show the differences in the mean $F_2$ values as a function of the preceding and following vowel at the onset, midpoint and offset of the English /ə/ and /æ/ produced by native and non-native speakers and the Japanese vowel /a/ produced by native Japanese speakers. These figures summarize the results of Experiment 2, 4b and the present experiment.

First of all, when Figures 11.1 and 11.2 are compared, the native speakers' schwa shows stronger carryover than anticipatory V-to-V effects. Significant

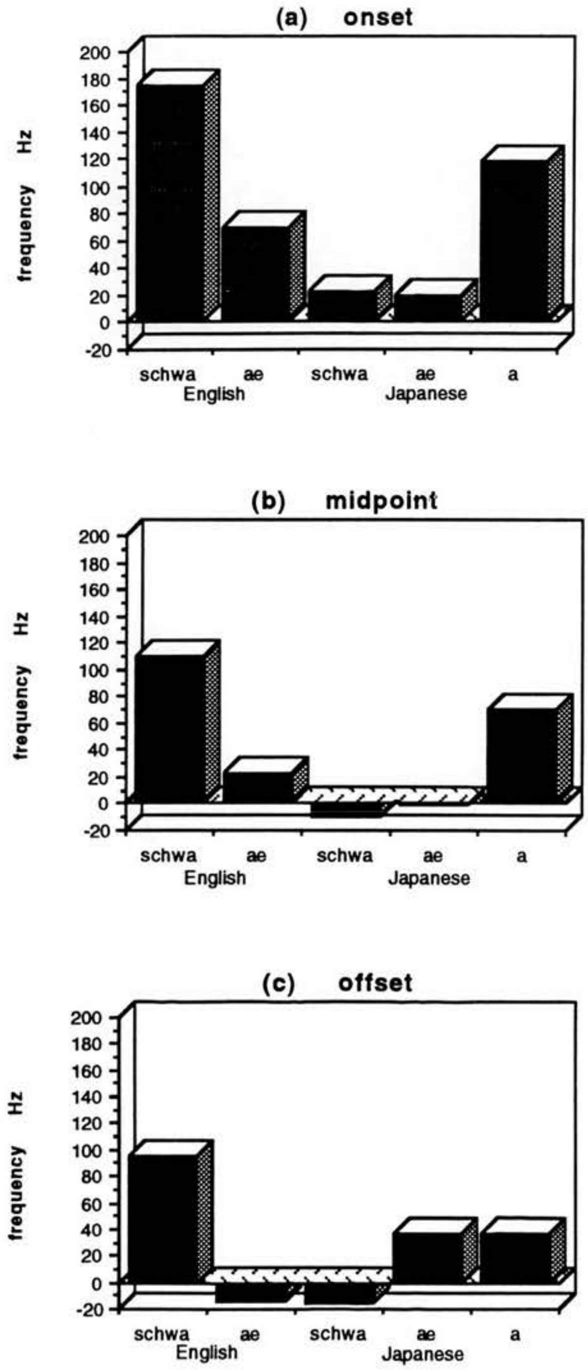



**Figure 11.1.** The differences in $F_2$ values as a function of the preceding vowels /ɪ/ (/i/) and /æ/ (/a/) for the English vowels /ə/ and /æ/ produced by English and Japanese speakers and the Japanese vowel /a/ produced by Japanese speakers.

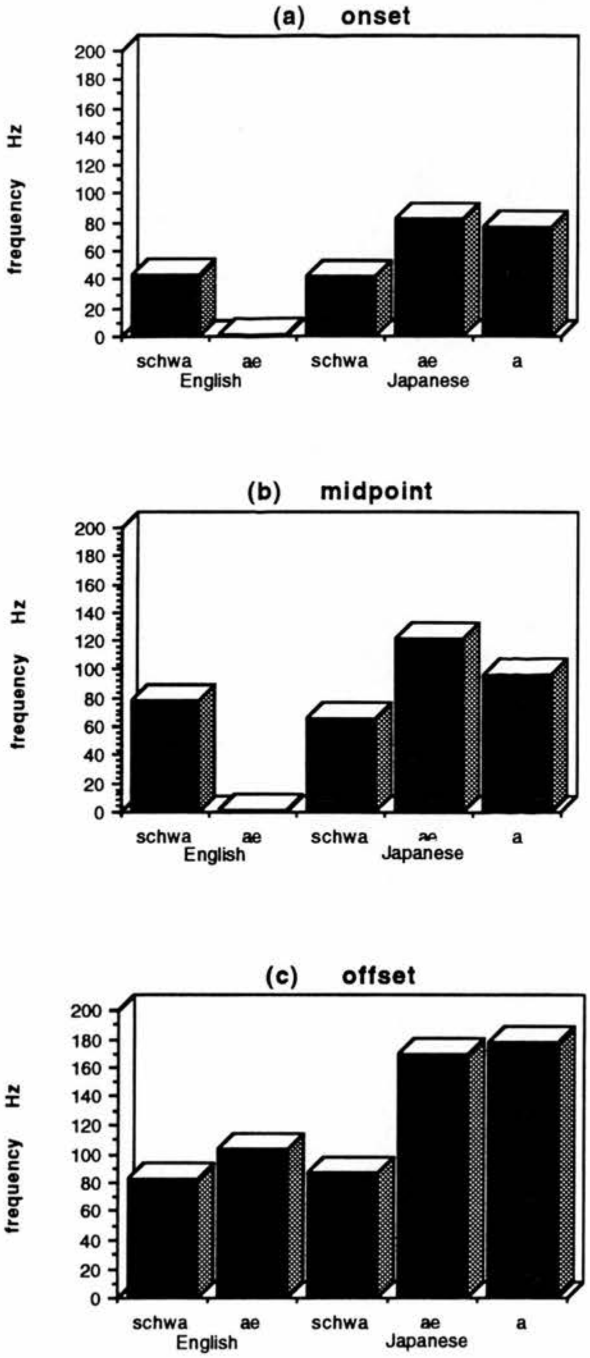



**Figure 11.2.** The differences in $F_2$ values as a function of the following vowels /ɪ/ (/i/) and /æ/ (/ə/ or /i/) for the English vowels /ə/ and /æ/ produced by English and Japanese speakers and the Japanese vowel /a/ produced by Japanese speakers.

differences are observed right through the segment for both carryover and anticipatory effects. For the full vowel /æ/, the differences are smaller compared to schwa for carryover effects and the effects diminish around the midpoint of the segment. For anticipatory effects, the difference is observed only at the offset of the vowel.

For the Japanese vowel /a/, anticipatory effects are greater in magnitude compared to carryover effects. Significant differences are observed right through the segment for both carryover and anticipatory effects. When the effects of the preceding vowel are compared between the native speakers' /ə/ and the Japanese /a/, the English /ə/ shows greater effects. The extent of V-to-V carryover effects observed on the Japanese vowel /a/ is intermediate in degree between that observed for the native speakers' /ə/ and /æ/. However, for anticipatory effects, the Japanese vowel /a/ shows the greatest effects among the three.

When the Japanese speakers' production of the English vowels are observed, the picture is somewhat distorted. First of all, contrary to the hypothesis above, Japanese speakers showed greater variability for the full vowel /æ/ than for /ə/. Secondly, the Japanese speakers hardly showed any carryover effects. Figure 11.3 shows the mean $F_2$ trajectories for the symmetric Vb_bV sequences with the English vowels /ə/ and /æ/ as the middle vowel produced by the Japanese subjects. The differences are observed throughout the vowels /ə/ and /æ/, but the differences are greater for the full vowel /æ/ than for /ə/.

Tables 11.3 through 11.7 show the differences in the mean $F_2$ values as a function of the contextual vowels for the English vowels /ə/ and /æ/ produced by native and non-native speakers and for the Japanese vowel /a/ produced by Japanese subjects in the symmetric and asymmetric contexts. The asterisk indicates that the difference has reached a statistically significant level of $p < 0.05$ (post hoc scheffe tests). The asterisk in a parenthesis means that the results of t-tests showed statistically significant difference at the level of $p < 0.05$.

When the mean $F_2$ values of the vowels /ə/ and /æ/ as a function of the preceding vowel are observed for each speaker (see Appendix A), two speakers, KN and MN, seem to show some systematic effects at the onset for both /ə/ and /æ/. These subjects are relatively fluent speakers of English. In Experiment 6, two fluent non-native speakers of English, MN and MT, showed stronger effects

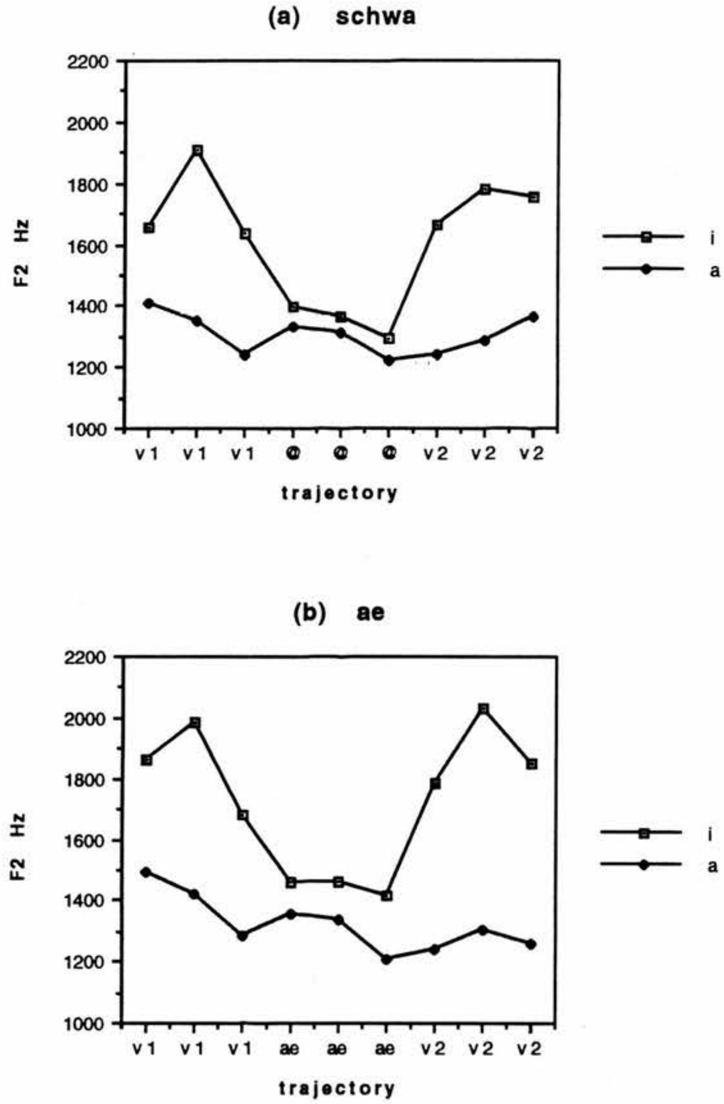



**Figure 11.3.** The mean $F_2$ trajectories in the symmetric Vb_bV sequence for the English /ə/ and /æ/ produced by Japanese speakers of English. The contextual vowels are /ɪ/ and /æ/ or /ə/.

| Symmetrical: Vb_bV | | | | |
|---|---|---|---|---|
| Speaker | vowel | onset | midpoint | offset |
| English | /ə/ | *215.3 | *180.4 | (*)161.0 |
| | /æ/ | (*)73.9 | 23.8 | 91.2 |
| Japanese | /ə/ | 64.8 | 55.9 | 72.0 |
| | /æ/ | (*)102.6 | (*)120.8 | *206 |
| | /a/ | *196 | (*)167.3 | *214.8 |

**Table 11.3.** The differences in $F_2$ value (Hz) as a function of the contextual vowels /ɪ/ (/i/) and /æ/ (/ə/ or /a/) for the schwa and the vowel /æ/ produced by English and Japanese speakers and the Japanese vowel /a/ produced by native Japanese speakers.

| Carryover: Vb_bɪ(/i/) | | | | |
|---|---|---|---|---|
| Speaker | vowel | onset | midpoint | offset |
| English | /ə/ | *188.9 | (*)106.6 | 96.7 |
| | /æ/ | (*)68.5 | 30.1 | 9.8 |
| Japanese | /ə/ | 20.3 | -18.4 | -36.1 |
| | /æ/ | 17.3 | 3.7 | 12.4 |
| | /a/ | 52.2 | 31.1 | 30.6 |

**Table 11.4.** The differences in $F_2$ value (Hz) as a function of the preceding vowels /ɪ/ (/i/) and /æ/ (/a/) for the schwa and the vowel /æ/ produced by English and Japanese speakers and the Japanese vowel /a/ produced by native Japanese speakers.

| Carryover: Vb_bæ(/ə/ or /a/) | | | | |
|---|---|---|---|---|
| Speaker | vowel | onset | midpoint | offset |
| English | /ə/ | (*)159.4 | (*)103.3 | (*)57.6 |
| | /æ/ | (*)70.3 | 13.7 | -33.5 |
| Japanese | /ə/ | 24.4 | -1.8 | 7.5 |
| | /æ/ | 22.0 | -5.8 | 61.2 |
| | /a/ | *187.6 | (*)111.0 | (*)52.9 |

**Table 11.5.** The differences in $F_2$ value (Hz) as a function of the preceding vowels /ɪ/ (/i/) and /æ/ (/a/) for the schwa and the vowel /æ/ produced by English and Japanese speakers and the Japanese vowel /a/ produced by native Japanese speakers.

| Anticipatory: ɪb_bV (ib_bV) | | | | |
|---|---|---|---|---|
| Speaker | vowel | onset | midpoint | offset |
| English | /ə/ | (*)55.9 | (*)77.1 | (*)103.4 |
| | /æ/ | 3.6 | 10.1 | (*)124.7 |
| Japanese | /ə/ | 40.4 | 57.7 | 64.5 |
| | /æ/ | (*)80.6 | (*)126.7 | (*)144.8 |
| | /a/ | 8.4 | 56.3 | (*)61.9 |

**Table 11.6.** The differences in $F_2$ value (Hz) as a function of the following vowels /ɪ/ (/i/) and /æ/ (/ə/ or /a/) for the schwa and the vowel /æ/ produced by English and Japanese speakers and the Japanese vowel /a/ produced by native Japanese speakers.

| Anticipatory: æb_bV (ab_bV) | | | | |
|---|---|---|---|---|
| Speaker | vowel | onset | midpoint | offset |
| English | /ə/ | 26.4 | 73.8 | (*)64.3 |
| | /æ/ | 5.4 | -6.3 | 81.4 |
| Japanese | /ə/ | 44.5 | 74.3 | (*)108.1 |
| | /æ/ | (*)85.3 | (*)117.1 | *193.6 |
| | /a/ | (*)143.8 | (*)136.2 | *194.2 |

**Table 11.7.** The differences in $F_2$ value (Hz) as a function of the following vowels /ɪ/ (/i/) and /æ/ (/ə/ or /a/) for the schwa and the vowel /æ/ produced by English and Japanese speakers and the Japanese vowel /a/ produced by native Japanese speakers.

of the contextual vowels at the onset than at the offset in the symmetric Vp_pV context; the differences of 109 Hz at the onset and 80 Hz at the offset for MN and the differences of 279 Hz at the onset and 205 Hz at the offset for MT. KN showed hardly any effects of the vocalic context in the Vp_pV context: the differences of -4.3 Hz at the onset and 24 Hz at the offset. In the same Japanese Vp_pV context, all the four native speakers of Japanese, including MN, invariably showed stronger anticipatory effects. MN showed the differences of 106 Hz at the onset and 209 Hz at the offset for his Japanese /a/ as a function of the contextual vowels /i/ and /a/.

This seems to suggest that L2 speakers in an advanced level may shift the relative strength in the directionality of coarticulation from R-to-L to L-to-R. This shift seems to be attainable only when these speakers have acquired the prosodic organization of the L2.

On the other hand, the Japanese speakers of English in the present experiment showed a unique coarticulatory pattern in the production of the English vowels. They showed the least contextual effects for /ə/ (see Figures 11.1 and 11.2) while the magnitudes of anticipatory V-to-V effects were similar between their Japanese /a/ and their English /æ/. They hardly showed any carryover effects for either /ə/ or /æ/. For the anticipatory effects on both /ə/ and /æ/, the Japanese speakers showed greater V-to-V effects when $V_1$ was /æ/ than /ɪ/ (compare Tables 11.6 and 11.7). This tendency also seems to be a transfer from the L1 coarticulatory strategy (see page 177 in Chapter 7).

KN who was rated as a relatively fluent speaker of English also behaved uniquely in the production of the English vowels in this experiment. He showed similar degrees of carryover effects for the English vowels /ə/ and /æ/, the differences of 70 Hz and 65 Hz as a function of the preceding vocalic context at the onset. These differences are somewhat smaller compared to the difference of 143 Hz observed at the onset of his Japanese vowel /a/. He showed rather substantial anticipatory effects for the vowel /æ/, the differences of 91 Hz, 177 Hz and 307 Hz at the onset, midpoint and offset of the vowel /æ/ as a function of the following vocalic context. However, he showed no systematic anticipatory effects for schwa at all. The extent of anticipatory effects observed on his Japanese vowel /a/ was less pronounced than that observed on his English /æ/, the differences of 34 Hz, 93 Hz and 115 Hz at the onset, midpoint and offset.

These results are interesting in that while fluent non-native speakers of English seem to have acquired an L2-like coarticulatory pattern for the combination of C-to-V and V-to-V effects, they show a rather unique pattern when V-to-V effects in the labial context /b/ is observed.

The pattern observed here is in a way a transfer from the L1 system in that the L1 coarticulatory pattern is over-emphasized in the interlanguage. That is, anticipatory effects are stronger than carryover effects in Japanese, but in the interlanguage, while anticipatory effects are markedly pronounced, carryover effects are nil. The Japanese speakers are also differentiating the coarticulatory patterns of /ə/ and /æ/. However, they seem to be doing this in a reverse manner from the L2 pattern, showing stronger contextual effects on the full vowel /æ/ than on the reduced vowel /ə/. KN showed the most pronounced trend of this. SK was the only subject who showed stronger anticipatory effects on /ə/ than on /æ/.

It is not quite clear why subjects like KN and MN who showed an L2-like coarticulatory pattern in Experiment 6 showed a deviant pattern. One possible reason for this deviant behaviour is that C-to-V and V-to-V coarticulation are essentially different and that the developmental processes in interlanguage reflect this. However, when C-to-V and V-to-V effects on schwa in Experiment 6 are closely observed, fluent non-native speakers of English exhibit patterns very much like those manifested by native speakers for both C-to-V and V-to-V effects. Tables 11.8 and 11.9 show the V-to-V effects in each consonantal context /p, t, k/ and the C-to-V effects in each vocalic context /ɪ, æ, u/ at the three points of measurement observed for the fluent group of speakers, KN, MN and MT in Experiment 6. The following similarities were observed between the native and fluent non-native speakers' schwa (Cf. Tables 3.6 and 3.7).

1. For the C-to-V coarticulation, anticipatory effects are stronger than carryover effects.
2. The C-to-V effects are stronger in the front vowel context than in the back vowel context.
3. The V-to-V effects are greatest in the context of /k/.
4. The V-to-V effects are small in the context of the alveolar /t/ (for KN and MN).
5. Carryover V-to-V effects are stronger than anticipatory effects in the labial /p/ context (for MN and MT).
6. Anticipatory effects are stronger than carryover effects in the velar context of /k/ (for MN and MT).

On the other hand, KN showed very little V-to-V effects in the labial context, but the native speaker DG also showed very little V-to-V effects in the labial context. MT showed rather strong V-to-V effects in the alveolar context. He generally showed very strong contextual effects that may be described as overshooting. In the vocalic context of /æ/, he showed the mean difference in $F_2$ of 1050 Hz as a function of the contextual consonants /p/ and /k/ at the vowel offset. With the limited number of subjects we may not draw any strong conclusion, particularly as idiosyncratic behaviour is observed in both C-to-V and V-to-V coarticulation (see page 68). However, the pattern observed in Tables 11.8 and 11.9 do not seem very deviant from the native speakers' patterns.

In Experiment 6, schwa in the indefinite article *a* was observed. Also, the sentences were relatively easy with familiar words. On the other hand, in the present

| subject | context | onset | midpoint | offset |
|---|---|---|---|---|
| KN | Vp_pV | -4.3 | 27.1 | 24 |
| | Vt_tV | -35.8 | -34.2 | -19.7 |
| | Vk_kV | *503.5 | *458.3 | *487.9 |
| MN | Vp_pV | 108.6 | 116.7 | 79.6 |
| | Vt_tV | -27.2 | 16.7 | 70.3 |
| | Vk_kV | *184.3 | *198.7 | *290.4 |
| MT | Vp_pV | 278.7 | 320.5 | 204.7 |
| | Vt_tV | *469.5 | 389.8 | 152.4 |
| | Vk_kV | 288.8 | *498.3 | *764.3 |

**Table 11.8.** The differences in the mean $F_2$ values of schwa as a function of the contextual vowels /ɪ/ and /u/ (/ɪ/ - /u/) at the different points in the trajectory for different consonantal contexts. The symbol * means that the difference was statistically significant ($p < 0.05$, post hoc scheffe tests).

| subject | context | onset | midpoint | offset |
|---|---|---|---|---|
| KN | iC_Ci | *525.1 | *544.5 | *582.0 |
| | aC_Ca | *512.9 | *478.2 | *550.9 |
| | uC_Cu | 120.4 | 170.5 | 198.7 |
| MN | iC_Ci | *296.1 | *319.3 | *440.3 |
| | aC_Ca | *274.9 | *301.0 | *403.1 |
| | uC_Cu | *319.3 | *350.4 | *360.8 |
| MT | iC_Ci | *538.2 | *611.4 | *830.6 |
| | aC_Ca | *862.0 | *715.3 | *1050.0 |
| | uC_Cu | 117.7 | *266.5 | *438.1 |

**Table 11.9.** The differences in the mean $F_2$ values of schwa as a function of the contextual consonants, /p/ and /k/ (/k/ - /p/) for the front vowel contexts and /p/ and /t/ (/t/ - /p/) for the back vowel context, at the different points in the trajectory. The symbol * means that the difference was statistically significant ($p < 0.05$, post hoc scheffe tests).

experiment lexical schwa was used. The content words used in the present experiment, such as *abysmal*, *fib* and *crib*, were relatively unfamiliar to the Japanese subjects. This may have affected their performance. The L2-like coarticulatory pattern of schwa observed in Experiment 6 may not immediately affect all the lexical items in the interlanguage. The acquired coarticulatory pattern may gradually spread from more familiar words to less familiar words. It is plausible that when L2 speakers are not very confident in their L2 performance, they shift back to an L1-like pattern.[3] This issue is related to systematic variability in interlanguage. The acquired coarticulatory pattern may still be variable and unstable and under certain circumstances may shift back to a more L1-like pattern. A further investigation is needed to see what circumstances favour a more L2-like pattern. It is also possible that the different performances observed are due to different subjects who participated in the two experiments. A more controlled experiment with the same subjects performing different tasks is necessary to address the question.

## 11.3 Conclusion

The results of the present experiment did not bear out the hypothesis presented at the beginning of this chapter. Contrary to the prediction, Japanese speakers of English showed less systematic contextual variability on /ə/ than on the full vowel /æ/. The pattern of V-to-V coarticulation across /b/ observed for the Japanese speakers' /ə/ and /æ/ may be characterized as a transfer or more adequately as an over-projection of the coarticulatory pattern of L1 onto the interlanguage. Their coarticulatory pattern of /ə/ was more deviant from the L1 pattern than that of /æ/. However, instead of approaching the L2 pattern they seem to have

[3]The difficulty of SLA studies involving the pronunciation of L2 is that it is quite impossible to ask L2 speakers to produce L2 nonsense words. On the other hand, to put certain sequence of sounds into natural sentences, some of the words may turn out to be unfamiliar to them. In the present experiment, most subjects did not know the words *abysmal*, *fib* and *crib* and they had to be told what some of the sentences meant before the recording. However, once they were told the meaning of the sentences and after some practice reading, they seemed to be quite comfortable and confident in producing the sentences. Their production sounded natural to the author.

shifted away from the L2 as well. On the other hand, the coarticulatory pattern of /æ/ was more L1-like. This seems to suggest that the Japanese subjects in the present study had some awareness of schwa being a unique and different vowel, but somehow failed to produce the correct coarticulatory pattern. Unfamiliarity of some of the words used in the experiment may account for at least part of their poor performance. Contextual variability, in the present case the lexical item in which coarticulation is observed, may affect the nature of interlanguage. This systematic variability may explain the discrepancy observed between the non-native speakers' performance in Experiments 6 and 7. However, a more controlled experiment is necessary to prove this.

# Chapter 12

# General Discussion and Conclusion

## 12.1 Discussion

The results of the present study suggest that the nature and extent of coarticulation observed on vowels in languages may be determined by various constraints defined at different levels of the linguistic representation, resulting in an interesting interaction between the phonetics and phonology of a language. Crosslinguistic varieties observed among languages may be accounted for by (1) the type of constraints observed in each language, and (2) the way different constraints interact with one another within a language. Some constraints may be universal in nature while others may be language specific. In this discussion, I will present my view on how these constraints may be organized to determine the nature of coarticulation in different languages. The description will be based on the observations made in the present study and those reported in previous studies on coarticulation. First of all, I will describe some universally motivated constraints on coarticulation. Then I will discuss some of the language specific constraints that may be arbitrary in nature and defined at the higher level of the linguistic representation. I will argue that the overall nature of the coarticulatory pattern of a language may be defined at this higher level. Lastly, I will show how the higher level constraints may interfere with physically motivated constraints on coarticulation.

According to Ohala (1983), 'the ultimate task of phonology is to discover the causes of the behaviour of speech sounds. To do this phonologists must refer to the way speech is created and used by humans, including how it is stored in the brain, retrieved, executed, perceived and used to facilitate social interaction among humans. The domain of phonology is therefore, mind, matter and manner.' The matter that humans use for speech across languages is limited in that the shape and size of the speech instrument, i.e., the vocal tract, are more or less the same across speakers of different languages. This implies that the same physical constraints (acoustic, articulatory, aerodynamic and perceptual) may apply on the organization of speech across languages.

For example, in the present study, the $F_1$ value of a vowel was observed to be higher in the labial /pVp/ context in both English and Japanese. The aerodynamic factors seem to explain this (Engstrand 1983; see also page 44). The acoustic cue of the CV transition seems to be most effective when there is less turbulence noise during the interval and this favours a lowered tongue position during /p/. Similarly, in the /NVN/ sequence, the acoustic cue for the place of articulation of the nasal consonant is more effectively perceived when the nasal coupling is reduced around the consonantal release or implosion. This results in the slightly raised velum during the vowel interval. Thus, in languages where the vowels are unspecified for [Nasality], the articulatory trajectory of the velum across the /NVN/ sequence is not a straight line interpolation, but it is characterized by the rising of the trajectory during the middle vowel portion. The $F_2$ trajectories across schwa in the symmetric VCəCV sequences observed in the present study were not level. Asymmetries in consonant loci between the CV and VC transitions have also been observed by Choi (1992) for Marshallese, and Kent & Moll (1972) and Munhall *et al.* (1991) for English. Similar asymmetries have also been observed for the Japanese data in the present study. These asymmetries may be partially explained by aerodynamic factors. In the CV transition, the pressure build-up behind the constriction results in the forward movement of the tongue at the consonant release while at the VC transition, air pressure has been equalized to atmosphere and thus no such movement is observed (Kent & Moll 1972). Thus, the $F_2$ values are slightly higher at the CV transition than at the

VC transition within each consonantal context.

At the periphery of the speech organization physical constraints such as described above seem to determine the fine detail of coarticulation. These physical constraints seem to be at the bottom of the constraint hierarchy. These constraints may be universal in nature and may affect most languages in the world as long as there is no conflicting constraint at a higher level of the hierarchy. See below for discussion on the cases where higher level constraints may interfere with lower level constraints.

Hardcastle (1982:p47) suggests that 'low-level constraints do undoubtedly exist and do account for many contiguous adjustment-type coarticulations, but cannot operate over extensive stretches of speech and so cannot account for the long-range type of coarticulatory phenomena. ...coarticulatory patterns are often language-specific, reflecting, and being constrained by, the phonological and syntactic rules of the language.' As I suggested in Chapter 2, the long-range type of coarticulatory phenomena that are generally described as phonological phenomena, e.g., vowel harmony and feature spreading, seem to be using the same phonetic materials as contiguous gestural overlap. Lip rounding, nasal coupling, tongue backing, etc. are all examples of features constrained by physiological factors. As each gesture needs to be executed over a certain period of time and gestures across different articulatory tiers naturally overlap in time, coarticulation is simply inevitable. However, the phonology of a language may control the extent in time and degree of coarticulation. Thus, certain coarticulatory phenomena, such as lip rounding and nasal coupling, may be stretched over a longer period of time in some languages due to phonological specifications.

At the higher level of the constraint hierarchy, there seem to be constraints that are defined by the prosodic organization of a language. I argued that in English, the prosodic unit defined by the alternations among full and reduced vowels, i.e., the stress foot, largely determines the overall coarticulatory pattern of the vowels. This constraint is defined by the phonology of the language, and its application surfaces as the phonetic underspecification of schwa and targeted full vowels in $F_2$. I also suggested that strong V-to-V carryover effects observed on schwa may be explained by this characteristic rhythm of English. If the rhythm of English is characterized by the basic beat from stressed to stressed vowel, it is intuitively correct that carryover effects are greater than anticipatory effects. The effects of a stressed vowel seem to decay across time towards the beginning of the next stressed vowel where a new pulse starts. The fact that non-native speakers of English shifted their coarticulatory pattern from an L1-like system to a more L2-like system where they showed extensive variability in $F_2$ as a function of contexts strongly supports the phonetic underspecification of schwa. Some of the non-native speakers also showed a shift in the strength of the directionality of V-to-V effects. This seems to suggest that the overall organization of speech has shifted from the L1 system to a more L2-like system.

In contrast to English, Japanese which does not have a contrast of full and reduced vowels manifests a completely different coarticulatory pattern. The characteristic features of the Japanese coarticulatory pattern are (1) generally great extent of V-to-V coarticulation, (2) strong anticipatory effects, and (3) long-range V-to-V effects when there is a sequence of syllables with identical vowels in them, in which case the vowel next to the affecting vowel seems to become transparent to the V-to-V effects. The extent of V-to-V carryover effects on the Japanese vowel /a/ was intermediate in degree between that observed for the English full vowel /æ/ and the reduced vowel /ə/. For anticipatory effects, however, the Japanese vowel /a/ showed the greatest effects among the three vowels. It is not quite clear why anticipatory effects are stronger than carryover effects in Japanese. The relationship between the Japanese coarticulatory pattern and the prosodic units that may define this pattern needs to be investigated in a future study to solve this question.

In Japanese there is a vowel weakening process called vowel devoicing. (See 5.5 for extensive discussion of this phenomenon.) The two high vowels /i/ and /u/ are particularly susceptible to this process. Jun & Beckman (1993, 1994) suggest that vowel devoicing is a result of gestural overlap. They consider that it is a universal vowel weakening process observed in all the languages. In English, schwa is the target of this process (See 3.2.2.) This process may hit the weakest vowels in the vowel inventories of the world languages.[1] Vowel devoicing is also

[1] Voicing, i.e., vibration of the vocal cords, has two physiological requirements: (1) the vocal cords must be lightly adducted, and (2) there must be sufficient air flowing through them. It is well recognized that it is difficult to maintain voicing during a stop. This is because air flowing

contextual assimilation. Devoiced vowels seem to be subsumed under the frication period of the surrounding segments. The treatment of vowel devoicing is an important theoretical issue. I would like to treat the vowel devoicing process as being on a different level from the vowel reduction process such as the one observed in English. There may be two different levels in vowel weakening process in English. One is phonological and the other is phonetic. Schwa is unspecified in the phonology of English and this underspecification surfaces to the phonetic level. The phonetic underspecification of schwa, however, is determined by the phonology of the language. This comprises the first level in the overall vowel weakening of the language. In the second level, the already weak segment, schwa is a target to the vowel devoicing process. It often gets further weakened, devoiced or deleted. This second level of weakening, though phonetic in nature, seems to be still under the constraints of the grammar of particular languages. Dalby (1986) suggests that the vowel devoicing of schwa is constrained by the prosody of the language. He found fewer instances of schwa devoicing or deletion where the syllable structure would have been affected. Kondo (forthcoming) also proposes that vowel devoicing in Japanese may be constrained by the mora structure.

Another example of interactions between different level constraints may be drawn from literature. It was suggested above that the nasal coupling does not favour the perception of the place of articulation of the consonants in the /NVN/ sequence. Thus, the velum height during the vowel in the /NVN/ sequence is partially determined by the acoustic constraint. However, this low level constraint may be interfered if the language choses to have a phonemic contrast of [Nasality] on vowels. The choice of [+nasal] and [−nasal] vowels in the phoneme inventory of a language is arbitrary and it already overrides the acoustic constraint mentioned

through the glottis accumulates in the oral cavity resulting in the equilibrium between the oral and subglottal pressure. When this happens, the air flow through the glottis diminishes and voicing stops. Also, a close vowel creates an appreciable oral back pressure of about 1cm $H_2O$ in comparison with open vowels. Although the magnitude of the oral back pressure is not very great. It reduces the transglottal pressure drop and could contribute to vowel devoicing in conjunction with other factors. Greenberg (1969) and Jaeger (1978) surveyed a total of 30 languages that allophonically devoice a subset of their vowel inventory. They found a strong tendency for high vowels to exhibit devoicing as opposed to low vowels. The inherently short durations of the high vowels also contribute to the tendency for deletion under gestural overlap (Han 1962a; Han 1962b).

above. Presumably, the velum height during the [+nasal] vowel is defined by the phonology of the language while the velum height during non-nasal vowels is defined by phonetic constraints. An interesting issue is, however, whether the velum height during the /NVN/ sequence will be affected by the presence or absence of the [+nasal] vowels in the language. Will the velum height in the /NVN/ sequence determined by the same phonetic constraints for languages with and without phonemic nasal vowels? Or will the velum height during the [−nasal] vowels in the nasal context comparably higher in the language with the phonemic contrast of [Nasality] in order to contrast with [+nasal] vowels? If the latter is the case, the phonology of the language interferes with the low level phonetic constraint in the fine detail of coarticulation. The results obtained by Clumeck (1976) support the latter case. He observed more extensive coarticulation for the anticipatory velic opening in English than French where there is a contrast of nasal and non-nasal vowels.

## 12.2 Conclusion

In the introduction of the present dissertation, the five hypotheses to be tested were presented. The first four hypotheses were born out by the results of the present study.

1. Schwa in British English seems to be phonetically unspecified in $F_2$. However, it was not clear if the $F_1$ value of schwa was targeted or not. It was suggested that schwa may retain the status of syllable nucleus by having some default vocalic height. In this sense, schwa may be specified in [Height] but unspecified in [Backness].

2. The native speakers of British English manifested a sharp contrast in the extent of V-to-V effects in $F_2$ between the full vowel /æ/ and the reduced vowel /ə/.[2]

[2]The difference in the magnitude of vowel variability between /æ/ and /ə/ was more pronounced for carryover effects.

3. Accent in Japanese did not affect the extent of V-to-V effects. Some speakers showed greater variability for accented vowels while others showed greater variability for unaccented vowels. In general no significant effects of accent were observed on the $F_2$ values of the Japanese vowels.

4. The extent of V-to-V effects observed for the Japanese vowel /a/ was intermediate in degree between that observed for the English full vowel /æ/ and the reduced vowel /ə/ for carryover effects. For anticipatory effects, the Japanese vowel /a/ showed the greatest effects among the three vowels.

The difference in accent system seems to affect the coarticulatory pattern of the two languages. This difference reflects the extent to which accent is correlated with the prosodic organization of a language. While accent plays an important role in organizing speech into prosodic units in English, accent plays only a minor role in the prosodic organization of Japanese. The presence of the contrast between targeted and targetless vowels defined by the stress foot in English and the absence of such contrast in Japanese seem to determine the coarticulatory patterns of the two languages. The stronger carryover effects in English as opposed to stronger anticipatory effects in Japanese may also be determined by the prosodic organization of the two languages.

The fifth hypothesis on the interlanguage of schwa was not born out by the present study. The Japanese speakers of English showed different degrees of V-to-V effects between the English /æ/ and /ə/, but they showed greater context dependent variability on /æ/ than on /ə/, which is the opposite pattern from the native speakers' pattern. In general, they showed a strong projection of the coarticulatory pattern of the L1 onto the interlanguage.

On the other hand, for the VCəCV sequences in Experiment 6 with the indefinite article *a*, a system which is intermediate in value between the L1 and L2 system was observed for the non-native speakers' production of schwa. Fluent speakers of English showed a coarticulatory pattern very much like that of the native speakers' schwa. In other words, their schwas showed systematic and large variability in $F_2$ as a function of contexts. On the other hand, non-fluent speakers of English showed a coarticulatory pattern similar to the L1 vowel /a/ in $F_2$ while they showed non-systematic variability in $F_1$. Some fluent speakers also showed a shift from the L1 pattern where anticipatory effects are stronger to the L2 pattern where carryover effects are stronger. The discrepancy observed between the results of Experiment 6 and 7 may be explained by the difference in the materials used or subjects participated in the two experiments. Further investigations are necessary to test this.

The results of the present study showed that some non-native speakers acquire an L2-like coarticulatory pattern. The results of Experiment 7 suggest that some Japanese speakers of English seem to have acquired the phonetic underspecification of schwa. In order to achieve this, they must have acquired the accent system of English. Higher level constraints such as accent system seem to determine the general pattern of coarticulation in languages. However, the coarticulatory patterns of languages seem to be determined by a complex interplay of various constraints at different levels as suggested above in the discussion, and the process of acquiring the correct coarticulatory pattern of the L2 may be far from an easy task involving the restructuring of the system from L1 to L2 at various levels of the linguistic representation.

# Bibliography


**Abercrombie, D.** 1964. A phonetician's view of verse structure. *Linguistics* 6.5–13.

**Akinaga, K.** 1985. *Kyootsuugo no akusento*, 70–116. Tokyo, Japan: The Japan Broadcasting Corporation (NHK).

**Amanuma, Y., K. Ootsubo, & O. Mizutani.** 1988. *Nihongo Onseigaku*. Tokyo, Japan: Kuroshio Shuppan.

**Anderson, S.R.** 1982. The analysis of French schwa. *Language* 58.535–573.

—— 1993. The significance of phonetic representation. In *Conference on the robustness of the language faculty: coping with incomplete information*, Utrecht, Holland. The Research Institute for Language and Speech (OTS).

**Archangeli, D.** 1988. Aspects of underspecification theory. *Phonology* 5.183–207.

——, & **D. Pulleyblank.** forthcoming. *The content and structure of phonological representations*. Cambridge, Mass.: MIT Press.

**Barry, M.C.** 1984. Connected speech: processes, motivations, models. *Cambridge Papers in Phonetics and Experimental Linguistics* 3.1–16.

**Bates, S.A.R.**, forthcoming. PhD thesis. University of Edinburgh.

**Beckman, M.E.** 1982. Segment duration and the 'mora' in Japanese. *Phonetica* 39.113–135.

—— 1986. *Stress and Non-stress Accent*. Dordrecht: Forris. Netherlands Phonetic Archives 7.

——, & A. SHOJI. 1984. Segment duration and the 'mora' in Japanese. *Phonetica* 39.113–135.

BELL-BERTI, F., & K.S. HARRIS. 1976. Some aspects of coarticulation. *Haskins Laboratories: Status Report on Speech Research* SR 45/46, 197-204.

——, & ——. 1979. Anticipatory coarticulation: some implications from a study of lip rounding. *Journal of the Acoustical Society of America* 65.1268–1270.

——, & ——. 1981. A temporal model of speech production. *Phonetica* 38.9–21.

——, & R.A. KRAKOW. 1991. Anticipatory velar lowering: A coproduction account. *Journal of the Acoustical Society of America* 90.112–123.

BENGUEREL, A.P., & H.A. COWAN. 1974. Coarticulation of upper lip protrusion in French. *Phonetica* 30.41–55.

BERGEM, D.R. VAN. 1993. Acoustic vowel reduction as a function of sentence accent, word stress and word class. *Speech Communication* 12.1–23.

——. 1994. A model of coarticulatory effects on schwa. *Speech Communication* 14.143–162.

BLADON, R.A.W., & A.H. AL-BAMERNI. 1976. Coarticulation resistance in English /l/. *Journal of Phonetics* 4.137–150.

BOHN, O., & J.E. FLEGE. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14.131–58.

BORDEN, G., & K.S. HARRIS. 1980. *Speech Science Primer*. Baltimore: Williams and Wilkins.

BOYCE, S.E., 1988. *The influence of phonological structure on articulatory organization in Turkish and in English: Vowel harmony and coarticulation*. Yale University dissertation.

—— 1990. Coarticulatory organization for lip-rounding in Turkish and in English. *Journal of the Acoustical Society of America* 88.2584–2595.

——, R.A. KRAKOW, & F. BELL-BERTI. 1991. Phonological underspecification and speech motor organization. *Haskins Laboratories Status Report on Speech Research* SR-105/106.141–152.

BROAD, J.D., & R.H. FERTIG. 1970. Formant-frequency trajectories in selected CVC-syllable nuclei. *Journal of the Acoustical Society of America* 46(6).1572–1582.

BROWMAN, C.P., & L. GOLDSTEIN. 1986. Towards an articulatory phonology. *Phonology Yearbook* 3.219–252.

——, & ——. 1990a. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics* 18.299–320.

——, & ——. 1990b. Tiers in articulatory phonology, with some implications for casual speech. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. by J. Kingston & M.E. Beckman. Cambridge: Cambridge University Press.

——, & ——. 1992a. Articulatory phonology: an overview. *Phonetica* 49.155–180.

——, & ——, 1992b. Targeted and targetless schwas. Paper presented at International Phonologietagung. Krems, Austria.

——, & ——. 1992c. Targetless schwa: an articulatory analysis. In *Papers in Laboratory Phonology II: Gesture, Segment, Prosody*, ed. by G.J. Docherty & D.R. Ladd, 26–56. Cambridge: Cambridge University Press.

BROWN, P., & C. FRASER. 1979. Speech as a marker of situation. In *Social Markers in Speech*, ed. by K. Scherer & H. Giles. Cambridge: Cambridge University Press.

CARAMAZZA, A., G. YENI-KOMSHIAN, E. ZURIF, & E. CARBONE. 1973. The acquisition of a new phonological contrast: in the case of stop consonants. *Journal of the Acoustical Society of America* 54.421–428.

**Carney, P.J., & K.L. Moll.** 1971. A cinefluorographic investigation of fricative consonant-vowel coarticulation. *Phonetica* 23.192–202.

**Catford, J.C.** 1964. Phonation types: the classification of some laryngeal components of speech production. In *In honour of Daniel Jones*, ed. by D. Abercrombie, D.B. Fry, P.A.D MacCarthy, N.C. Scott, & J.L.M. Trim. London: Longman.

**Choi, J.D.** 1990. Kabardian vowels revisited. *UCLA Working Papers in Phonetics* 74.1–15.

—— 1992. Phonetic underspecification and target-interpolation: An acoustic study of marshallese vowel allophony. *UCLA Working Papers in Phonetics* 82.1–133.

——, & **P. Keating.** 1991. Vowel-to-vowel coarticulation in three Slavic languages. *UCLA Working Papers in Phonetics* 78.78–86.

**Chomsky, N.** 1976. *Reflections on Language*. London: Temple Smith.

—— 1981a. *Lectures on Government and Binding*. Dordrecht: Foris.

—— 1981b. Principles and parameters in syntactic theory. In *Explanation in Linguistics: the Logical Problem of Language Acquisition*, ed. by N. Hornstein & D. Lightfoot. London: Longman.

**Clark, J., & C. Yallop.** 1990. *An Introduction to Phonetics and Phonology*. Oxford: Blackwell.

**Clements, G.N.** 1985. The geometry of phonological features. *Phonology Yearbook* 2.225–252.

—— 1987. Towards a substantive theory of feature specification. In *Paper presented at the UCLA Symposium on Segment Structure*.

—— 1992. Phonological primes: gestures or features? *Phonetica* 49.222–234.

**Clumeck, H.** 1976. Patterns of soft palate movements in six languages. *Journal of Phonetics* 4.337–351.

COHN, A.C. 1993. Nasalisation in English: phonology or phonetics. *Phonology* 10.43–81.

COMRIE, B. 1984. *Language Universals and Linguistic Typology*. Oxford: Basil Blackwell.

CORDER, S.P. 1976. The study of interlanguage. In *Proceedings of the Fourth International Conference of Applied Linguistics*, Munich, Hochschulverlag.

—— 1977. Language continual and interlanguage hypothesis. In *Proceedings of the Fifth Neuchatel Colloquium*.

—— 1978. Language-learner language. In *Understanding Second and Foreign Language Learning: Issues and Approaches*, ed. by J. Richards. Rowley, Mass.: Newbury House.

CROFT, W. 1990. *Typology and Universals*. Cambridge: Cambridge University Press.

CROTHERS, J. 1978. Typology and universals of vowel systems. In *Universals of human language*, ed. by J.N. Greenberg, 93–152. Stanford: Stanford University Press.

CRYSTAL, D. 1985. *A Dictionary of Linguistics and Phonetics*. Oxford: Blackwell.

DALBY, J. 1986. *Phonetic structure of fast speech in American English*. Bloomington, Indiana: Indiana University Linguistics Club.

DANILOFF, R., & K. MOLL. 1968. Coarticulation of lip rounding. *Journal of Speech and Hearing Research* 11.707–721.

DE JONG, K., M.E. BECKMAN, & J. EDWARDS. 1993. The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2,3).197–212.

DELATTRE, P. 1969. An acoustic and articulatory study of vowel reduction in four languages. *IRAL* VII/4.295–325.

**Dickerson**, L. 1975. The learner's interlanguage as a system of variable rules. *TESOL Quarterly* 9.401–7.

**Disner**, S.F. 1983. Vowel quality: the relation between universal and languages specific factors. *UCLA Working Papers in Phonetics* 58.

**Docherty**, G.J., 1989. *An experimental phonetic study of the timing of voicing in English obstruents*. University of Edinburgh dissertation.

**Durand**, J. 1990. *Generative and Non-linear Phonology*. London: Longman.

**Ellis**, R. 1992. *Second Language Acquisition & Language Pedagogy*. Clevedon: Multilingual Matters Ltd.

—— 1994. *The Study of Second Language Acquisition*. Oxford: Oxford University Press.

**Engstrand**, O. 1983. Articulatory coordination in selected VCV utterances: a means-end view. *Reports from Uppsala University, Department of Linguistics* 10.

**Engstrand**, O. 1988. Articulatory correlates of stress and speaking rate in swedish vcv utterances. *Journal of the Acoustical Society of America* 83.1863–1875.

**Engstrand**, O. 1992. Systematicity of phonetic variation in natural discourse. *Speech Communication* 11.337–346.

**Fant**, G. 1960. *Acoustic Theory of Speech Production*. Hague: Mouton.

—— 1973. *Speech sounds and features*. Cambridge, MA: The MIT press.

**Farnetani**, E. 1990. V-C-V lingual coarticulation and its spatiotemporal domain. In *Speech Production and Speech Modeling*, ed. by W.J. Hardcastle & A. Marchal, 93–130. Dordrecht: Kluwer Academic Press.

——, & D. **Recasens**. 1993. Anticipatory consonant-to-vowel coarticulation in the production of VCV sequences in Italian. *Language and Speech* 36(2,3).279–302.

FLEGE, J.E. 1980. Phonetic approximation in second language acquisition. *Language Learning* 30.117–134.

—— 1981. The phonological basis of foreign accent: A hypothesis. *TESOL Quarterly* 15.443–455.

—— 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76.672–690.

—— 1988. Effects of speaking rate on tongue position and velocity of movement in vowel production. *Journal of the Acoustical Society of America* 84(3).901–906.

—— 1992. Speech learning in a second language. In *Phonological development: models, research, implications*, ed. by C.A. Ferguson, L. Menn, & C. Stoel-Gammon. Maryland: York Press.

——, & R. HAMMOND. 1982. Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition* 5.1–17.

——, & J. HILLENBRAND. 1984. Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America* 76(3).708–721.

——, & R. PORT. 1981. Cross-language phonetic interference: Arabic to English. *Language and Speech* 24.125–146.

FOWLER, C.A. 1981. Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 24.127–139.

FRY, D.B. 1958. Experiments in the perception of stress. *Language and Speech* 1.126–152.

GAY, T. 1974. A cinefluorographic study of vowel production. *Journal of Phonetics* 2.255–266.

—— 1978. Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America* 63(1).223–230.

GHAZELI, S., 1977. *Back consonants and backing coarticulation in Arabic*. University of Texas dissertation.

GIBBON, F., W. HARDCASTLE, & K. NICOLAIDIS. 1993. Temporal and spatial aspects of lingual coarticulation in /kl/ sequences: a cross-linguistic investigation. *Language and Speech* 36(2,3).261–277.

GIMSON, A.C. 1980. *An Introduction to the Pronunciation of English*. London: Edward Arnold.

GOLDSMITH, J. 1979. *Autosegmental Phonology*. New York: Garland Publishing.

GREENBERG, H. 1966. *Universals of Language (2nd ed.)*. Cambridge, Mass.: MIT Press.

—— 1969. Some methods of dynamic comparison in linguistics. In *Substance and structure of language*, ed. by J. Puhvel, 147–204. Los Angeles: Center for Research in Languages and Linguistics.

HABIS, A., forthcoming. PhD thesis. University of Edinburgh.

HAN, M.S., 1962a. *Japanese phonology: an analysis based on sound spectrograms*. University of Texas dissertation.

—— 1962b. Unvoicing of vowels in Japanese. *The Study of Sounds* 10.81–100.

HARDCASTLE, W.J. 1982. Constraints on coarticulatory processes. In *Linguistic Controversies: Essays in Linguistic Theory and Practice in Honour of F.R. Palmer*, ed. by D. Crystal, 33–49. London: Edward Arnold.

HAWKINS, J. 1983. *Word Order Universals*. New York: Academic Press.

HENKE, W, 1966. *Dynamic articulatory model of speech production using computer simulation*. MIT dissertation.

HIRAYAMA, T. 1985. *Zen Nihon no hatuon to akusento*, 37–69. Tokyo, Japan: The Japan Broadcasting Corporation (NHK).

HIROSE, H., Z. SHIMADA, & O. FUJIMURA. 1970. An electromyographic study of the activity of the laryngeal muscles during speech utterances. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 5.9–25.

HOCKETT, C.F. 1955. A manual of phonology. In *International journal of American linguistics (Memoir II)*. baltimore: Waverly Press.

HOMMA, Y. 1981. Durational relationship between Japanese stops and vowels. *Journal of Phonetics* 9.273–281.

HUFFMAN, M.K. 1986. Patterns of coarticulation in English. *UCLA Working Papers in Phonetics* 63.26–47.

HYAMS, N., 1983. *The acquisition of parameterized grammars*. City University of New York dissertation.

IMAISHI, M., K. SATO, J. MIWA, N. YOSHIDA, K. OHASHI, & M. KATO. 1984. Nihongo hoogen onsei no supekutoru bunseki shiryooshuu. In *Jyoohooka shakai ni okeru gengo no hyoojunka*. The Ministry of Education.

JAEGER, J.J. 1978. Speech aerodynamics and phonological universals. *Proceedings of the Annual Metting of the Berkeley Linguistics Society* 4.311–329.

JOHNSON, K., E. FLEMMING, & R. WRIGHT. 1993. The hyperspace effect: phonetic targets are hyperarticulated. *Language* 69(3).505–528.

JUN, SUN-AH. 1993. Asymmetry of prosodic effects on the glottal gesture in Korean. In *Paper presented at the Fourth Conference on Laboratory Phonology*, Oxford.

——, & M.E. BECKMAN. 1993. A gestural-overlap analysis of vowel devoicing in Japanese and Korean. In *Paper presented at the 1993 Annual Meeting of the Linguistic Society of America*, Los Angeles, California.

——, & ——. 1994. Distribution of devoiced high vowels in Korean. In *Proceedings 1994 International Conference on Spoken Language Processing*, Yokohama, Japan.

KAGAYA, R. 1974. A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2.161–180.

KAWAKAMI, S. 1977. *Nihongo Onsee Gaisetsu*. Tokyo, Japan: Oohuusha.

KEATING, P.A. 1985. CV phonology, experimental phonetics, and coarticulation. *UCLA Working Papers in Phonetics* 62.1–13.

—— 1988. Underspecification in phonetics. *Phonology* 5.275–292.

—— 1990. The window model of coarticulation. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. by J. Kingston & M.E. Beckman, 451–470. Cambridge: Cambridge University Press.

——, & M.K. HUFFMAN. 1984. Vowel variation in Japanese. *Phonetica* 59.50–61.

——, & A. LAHIRI. 1993. Fronted velars, palatalized velars and palatals. *Phonetica* 50.73–101.

KELLERMAN, E. 1983. Now you see it, now you don't. In *Language Transfer in Language Learning*, ed. by S. Gass & Selinker. L. Rowley, Mass.: Newbury House.

KENT, R., P. CARNEY, & L. SEVEREID. 1974. Velar movement and timing: Evaluation of a model for binary control. *Journal of Speech and Hearing Research* 17.470–488.

KENT, R.D., & K.L. MOLL. 1972. Cinefluorographic analyses of selected lingual consonants. *Journal of Speech and Hearing Research* 15.453–473.

KENYON, J.S. 1946. *American Pronunciation*. Ann Arbor, Michigan: George Wahr, 9th edition.

KINDAICHI, H. 1985. *Kyootsuugo no hatsuon to akusento*, 5–36. Tokyo, Japan: The Japan Broadcasting Corporation (NHK).

KIRITANI, S., H. ITOH, H. HIROSE, & M. SAWASHIMA. 1977. Coordination of the consonant and vowel articulations - X-ray microbeam study on Japanese and English. *Annual Bulletin of Research Institute for Logopedics and Phoniatrics* 11.11–21.

KONDO, M. 1993. The effect of blocking factors and constraints on consecutive vowel devoicing in standard Japanese. In *Poster presented at Labphon 4*, Oxford.

——, forthcoming. PhD thesis. University of Edinburgh.

KONDO, Y., 1989. L1 and L2 interference on the vowel formants of the English vowels [ʊ] and [u] and the Japanese vowels [ɯ] and [ɯɯ]. Master's thesis, University of Edinburgh.

KOOPMANS-VAN BEINUM, F. 1994. What's in a schwa? *Phonetica* 51.68–79.

KOZHEVNIKOV, V., & L. CHISTOVICH. 1965. *Speech: articulation and perception*. Washington DC: Joint Publications Research Service.

KRASHEN, S. 1981. *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

KUBOZONO, H., 1987. *The organization of Japanese prosody*. University of Edinburgh dissertation.

KUEHN, D., & K. MOLL. 1972. Perceptual effects of forward coarticulation. *Journal of Speech and Hearing Research* 15.654–664.

KUWABARA, H. 1972. An approach to normalization of coarticulation effects for vowels in connected speech. *Journal of the Acoustical Society of America* 77(2).686–694.

LABOV, W. 1970. The study of language in its social context. *Studium Generale* 23.3–87.

LADEFOGED, P. 1962. *Elements of Acoustic Phonetics*. Chicago: The University of Chicago Press.

——, & M. HALLE. 1988. Some major features of the International Phonetic Alphabet. *Language* 64.577–582.

——, & I. MADDIESON. 1990. Vowels of the world's languages. *Journal of Phonetics* 18.93–122.

LAVER, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

LEHISTE, I. 1970. *Suprasegmentals*. Cambridge, Ma.: MIT press.

——, & G.E. PETERSON. 1959. Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America* 31.428–435.

LEOPOLD, W. 1947. *Speech development of a bilingual child: Sound learning in the first two years*, volume 2. Evanston: Northwestern University Press.

LINDBLOM, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35.1773–1781.

LINDBLOM, B. 1983. Economy of speech gestures. In *The Production of Speech*, ed. by P.F. MacNeilage, 217–245. New York: Springer Verlag.

LINDBLOM, B. 1990. Explaining phonetic variation: a sketch of the H and H theory. In *Speech Production and Speech Modelling*, ed. by W.J. Hardcastle & A. Marchal. Netherlands: Kluwer Academic Publishers.

LINDBLOM, B, & S.-J. MOON. 1988. Formant undershoot in clear and citation form speech. *Phonetic Experimental Research, Institute of Linguistics, University of Stockholm (PERILUS)* 8.21–33.

LOVIN, J.B. 1976. Pitch accent and vowel devoicing in Japanese: A preliminary study. *Annual Bulletin of Research Institute for Logopedics and Phoniatrics* 10.113–125.

MAGEN, H. 1984. Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America* 75, Suppl..1:S11.

——, 1989. *An acoustic study of vowel-to-vowel coarticulation in English.* University of Yale dissertation.

MAJOR, R.C. 1977. Phonological differentiation of a bilingual child. *Ohio State University Working Papers in Linguistics* 22.88–122.

—— 1986. Paragogue and degree of foreign accent in Brazilian English. *Second Language Reserach* 2.53–71.

MAKHOUL, J., & L. COSELL. 1976. LPCW: An LPC vocoder with linear predictive spectral warping. In *Proceedings of ICASSP*, 466–469, Philadelphia.

MANUEL, S. 1990. The role of contrast in limiting vowel-to-vowel coarticulation in different languages. *Journal of the Acoustical Society of America* 88(3).1286–1298.

MANUEL, S., & R.A. KRAKOW. 1984. Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research* SR-77/78.69–78.

MARTIN, S.E., 1952. Morphophonemics of Standard Colloquial Japanese. Supplement to Language, Language Dissertation. No.47.

MCCAWLEY, J.D. 1968. *The Phonological Component of a Grammar of Japanese*. Hague: Mouton.

MCCLEAN, M. 1973. Forward coarticulation of velar movement at marked junctural boundaries. *Journal of Speech and Hearing Research* 16.286–296.

MENZERATH, P., & A. LACERDA. 1933. *Koartikulation, Steuerung und Lautabgrenzung: Eine experimentelle Untersuchung*. Berlin and Bonn: Ferd. Dü mmler.

MOLL, K.L., & R.G. DANILOFF. 1970. Investigation of the timing of velar movements during speech. *Journal of the Acoustical Society of America* 50.678–684.

**Moon, S.-J.**, 1991. *An acoustic and perceptual study of undershoot in clear and citation-form speech*. Austin: University of Texas dissertation.

——, & **B. Lindblom**. 1989. Formant undershoot in clear and citation-form speech: A second progress report. *Speech Transmission Laboratory Quarterly Progress and Status Report* 2.121–123.

**Munhall, K.**, & **A. Löfqvist**. 1992. Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics* 20.111–126.

**Munhall, K.G.**, **D.J. Ostry**, & **J.R. Flanagan**. 1991. Coordinated spaces in speech planning. *Journal of Phonetics* 19.293–307.

**Nolan, F.J.** 1985. Idiosyncrasy in coarticulatory strategies. *Cambridge Papers in Phonetics and Experimental Linguistics* 4.

**Nord, L.** 1974. Vowel reduction - centralization or contextual assimilation? In *Proceedings of the Speech Communication Seminar*, ed. by G. Fant, volume 2, 149–154.

—— 1986. Acoustic studies of vowel reduction in Swedish. *KTH-QPSR, Stockholm* 4.19–36.

**Norris, D.**, & **A. Cutler**. 1985. Juncture detection. *Linguistics* 23.689–705.

**Odlin, T.** 1989. *Language Transfer*. Cambridge: Cambridge University Press.

**Ohala, J.J.** 1983. The origin of sound patterns in vocal tract constraints. In *The Production of Speech*, ed. by P.F. MacNeilage. New York: Springer-Verlag.

—— 1993. Coarticulation and phonology. *Language and Speech* 36(2,3).155–170.

**Ohde, R.N.**, & **D.J. Sharf**. 1975. Coarticulatory effects of voiced stops on the reduction of acoustic vowel targets. *Journal of the Acoustical Society of America* 58.923–927.

——, & ——. 1977. Order effect of acoustic segments of VC and CV syllables on stop and vowel identification. *Journal of Speech and Hearing Research* 20.543–554.

**Öhman**, S.E.G. 1966. Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39.151–168.

**Ostreicher**, H.J., & D.J. **Sharf**. 1976. Effects of coarticulation on the identification of deleted consonant and vowel sounds. *Journal of Phonetics* 4.285–301.

**Pierrehumbert**, J., & M.E. **Beckman**. 1988. *Japanese Tone Structure*. Cambridge, Massachusetts: The MIT Press.

**Port**, R.F., S. **Ali-Ani**, & S. **Maeda**. 1980. Temporal compensation and universal phonetics. *Phonetica* 18.235–252.

——, & F **Mitleb**. 1981. Phonetic and phonological manifestations of the voicing contrast in Arabic-accented English. *Journal of Applied Psycholinguistics* 1.

**Prince**, A., & P. **Smolensky**. 1993. *Optimality Theory: constraint interaction in generative grammar*. Cambridge, Mass.: MIT. Ms., Rutgers University, New Brunswick, N.J., and University of Colorado, Boulder.

**Pulleyblank**, D. 1988. Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19.233–270.

**Recasens**, D. 1984. V-to-V coarticulation in Catalan VCV sequences: an articulatory and acoustical study. *Journal of Phonetics* 12.61–73.

—— 1986. *Fonètica experimental del català oriental central*. Barcelona: Publicacions de l'Abadia de Montserrat.

—— 1991. An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics* 19.177–192.

**Sawashima**, M. 1971. Devoicing of vowels. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 5.7–13.

——, H. **Hirose**, & H. **Yoshioka**. 1978. Adductor (PCA) and abduction (INT) muscles of the larynx in voiceless sound production. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 12.53–60.

**SELINKER, K.** 1972. Interlanguage. *International Review of Applied Linguistics* 10.209–31.

**SELKIRK, E.** 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11.563–605.

**SHARF, D.J., & R.C. BEITER.** 1974. Identification of consonants from formant transitions presented forward and backward. *Language and Speech* 17.110–118.

——, & **T. HEMEYER.** 1972. Identification of place of consonant articulation from formant transitions. *Journal of the Acoustical Society of America* 51.652–658.

——, & **R.N. OHDE.** 1981. Physiologic, acoustic, and perceptual aspects of coarticulation: implications for the remediation of articulatory disorders. In *Speech and Language: Advances in Basic Research and Practice*, volume 5, 153–247. Academic Press.

**SHOCKEY, L.** 1973. Phonetic and phonological properties of connected speech. *Ohio State Working Papers in Linguistics* 17.

**SMITH, C. L.** 1991. The timing of vowel and consonant gestures in Italian and Japanese. In *Proceedings of the XIIth International Congress of Phonetic Sciences*, Aix en-Provence, France.

**SOLI, S.D.** 1981. Second formants in fricatives: acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America* 70.976–984.

**STERIADE, D.** 1987. Redundant values. In *Papers from the parasession on autosegmental and metrical phonology*, ed. by A. Bosch, B. Need, & E. Schiller. Chicago: Chicago Linguistic Society.

**STEVENS, K.N.** 1990. Some factors influencing the precision required for articulatory targets: Comments on Keating's paper. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. by J. Kingston & M.E. Beckman, 471–475. Cambridge: Cambridge University Press.

——, & A.S. House. 1963. Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research* 6.111–128.

Takeda, K., & H. Kuwabara. 1987. Analysis and prediction of devocalizing phenomena. In *Proceedings of Acoustic Society of Japan*, 105–106.

Tarone, E. 1983. On the variability of interlanguage systems. *Applied Linguistics* 4.143–63.

Vassière, J. 1983. Prediction of articulatory movement of the velum from phonetic input. *Ms. Bell Laboratories* .

Weitzman, R.S., M. Sawashima, H. Hirose, & T. Ushijima. 1976. Devoiced vowel and whispered vowels in Japanese. *Annual Bulletin of Research Institute of Logopedics and Phoniatrics* 10.61–79.

Wells, J. 1982a. *Accents of English: An Introduction*. Cambridge: Cambridge University Press.

—— 1982b. *Accents of English: Beyond the British Isles*. Cambridge: Cambridge University Press.

Wenk, B. 1986. Crosslinguistic influence in second language phonology: speech rhythms. In *Crosslinguistic Influence in Second Language Acquisition*, ed. by E. Kellerman & S. Smith. Oxford: Pergamon.

Whalen, D.H. 1990. Coarticulation is largely planned. *Journal of Phonetics* 18.3–35.

Whitney, W. 1885. *The roots, verb-forms, and primary derivatives of the Sanskrit language*. Leipzig: Breitkopf and Härtel.

—— 1889. *Sanskrit grammar*. Cambridge, Mass.: Harvard University Press. Reprinted 1931.

Williams, L. 1980. Phonetic variation as a function of second-language learning. In *Child Phonology, vol.2, Perception*, ed. by G. Yeni-Komshian, J. Kavanegh, & C. Ferguson. New York: Academic Press.

WINITZ, H., M.E. SCHEIB, & J.A. REEDS. 1972. Identification of stops and vowels for the burst portion of the /p, t, k/ isolated from conversational speech. *Journal of the Acoustical Society of America* 51.1309–1317.

YENI-KOMSHIAN, G.H., & S.D. SOLI. 1981. Recognition of vowels from information in fricatives: perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America* 70.966–975.

YOSHIDA, N., & Y. SAGISAKA. 1990. Factor analysis of vowel devoicing in Japanese. *ATR Technical Report TR-I-0159* .

YOSHIOKA, H. 1981. Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38.236–2241.

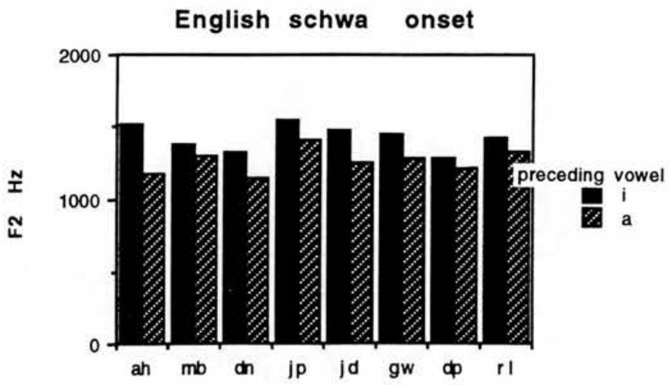
ZOBL, H. 1984. Cross-language generalizations and the contrastive dimension of the interlanguage hypothesis. In *Interlanguage*, ed. by A. Davies, C. Criper, & A. Howatt. Edinburgh: Edinburgh University Press.

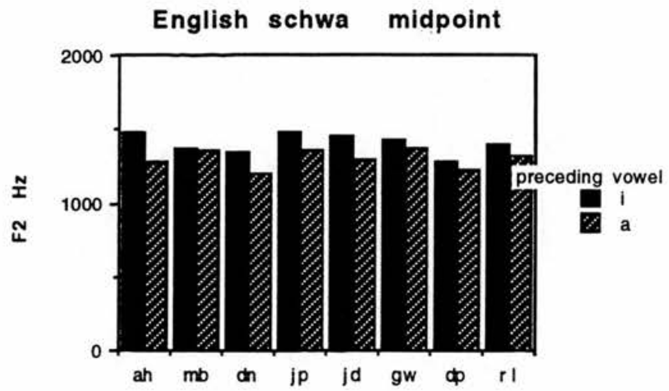
ZSIGA, E. 1993. An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In *the Fourth Conference on Laboratory Phonology*, Oxford.

ZWICKY, A. M. 1972. On casual speech. In *Papers from the 8th Regional Meeting of the Chicago Linguistic Society*, 607–615.

# Appendix A

This appendix contains figures that show the mean $F_2$ values as a function of the preceding and the following vowel at the onset, midpoint and offset of the English vowels /ə/, /æ/ and the Japanese vowel /a/ produced by each (native and non-native) speaker.


English schwa onset
2000
1000
0
F2 Hz
ah
mb
dn
jp
jd
gw
dp
rl
preceding vowel
i
a


English schwa midpoint
2000
1000
0
F2 Hz
ah
mb
dn
jp
jd
gw
dp
rl
preceding vowel
i
a

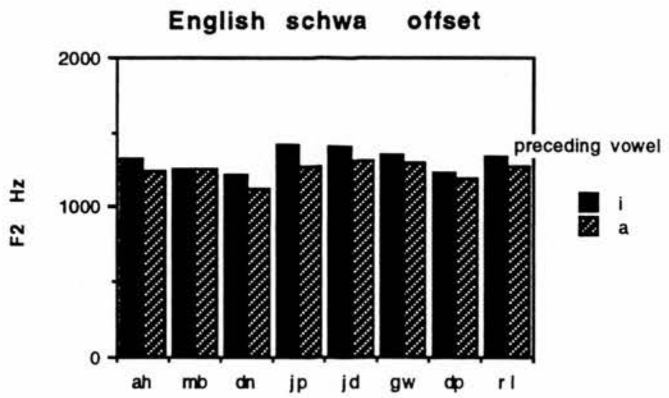

English schwa offset
2000
1000
0
F2 Hz
ah
mb
dn
jp
jd
gw
dp
rl
preceding vowel
i
a

Figure A.1.

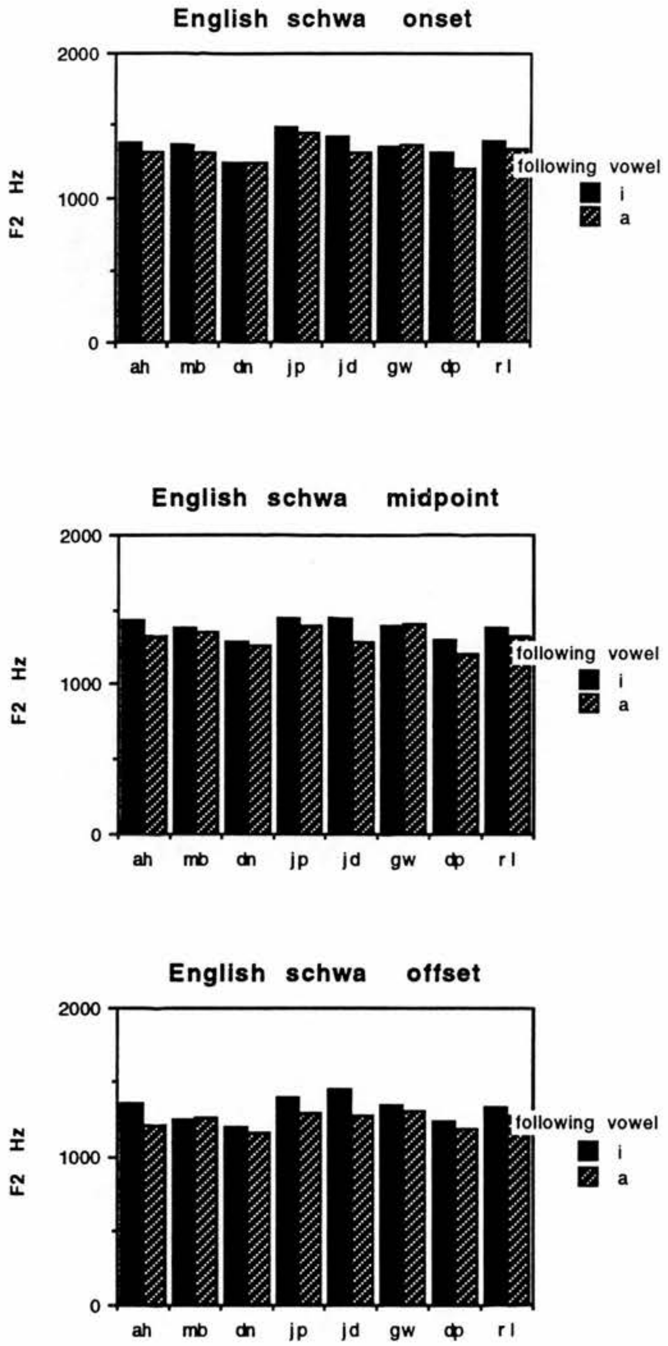

English schwa onset
2000
1000
0
F2 Hz
ah
mb
dn
jp
jd
gw
dp
rl
following vowel
i
a
English schwa midpoint
2000
1000
0
F2 Hz
ah
mb
dn
jp
jd
gw
dp
rl
following vowel
i
a
English schwa offset
2000
1000
0
F2 Hz
ah
mb
dn
jp
jd
gw
dp
rl
following vowel
i
a

Figure A.2.

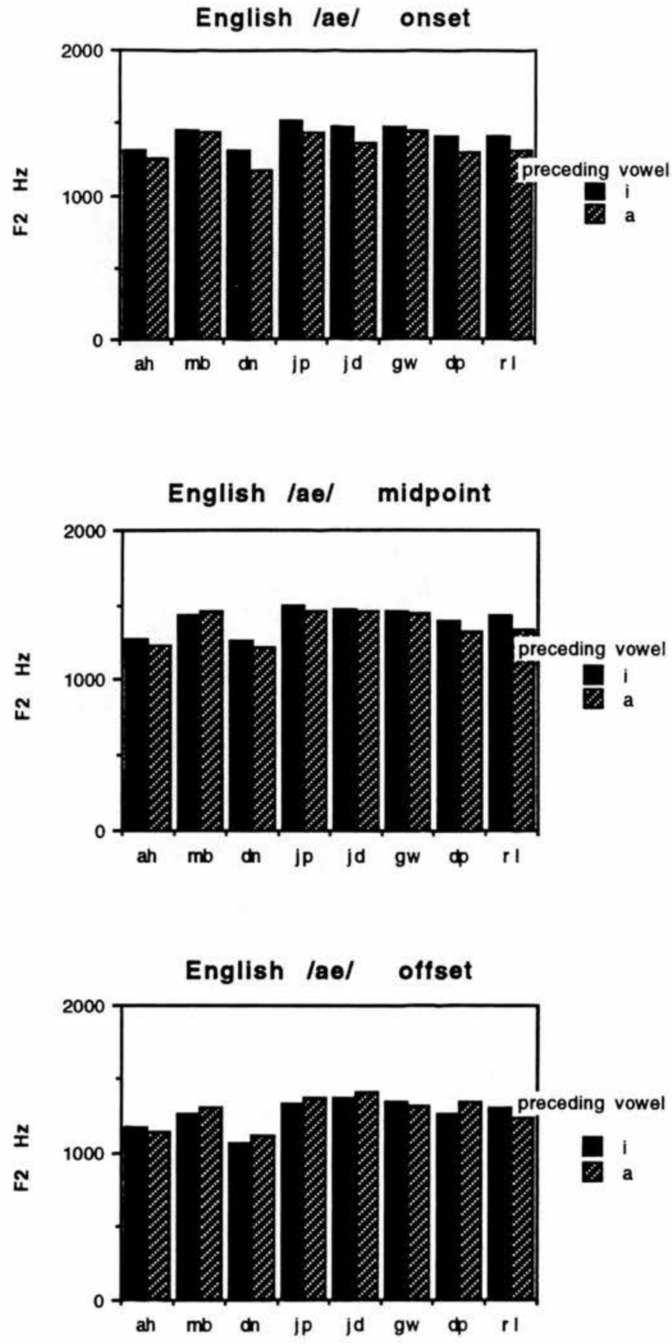

English /ae/ onset
2000
1000
0
F2 Hz
preceding vowel
i
a
ah mb dn jp jd gw dp rl
English /ae/ midpoint
2000
1000
0
F2 Hz
preceding vowel
i
a
ah mb dn jp jd gw dp rl
English /ae/ offset
2000
1000
0
F2 Hz
preceding vowel
i
a
ah mb dn jp jd gw dp rl

Figure A.3.

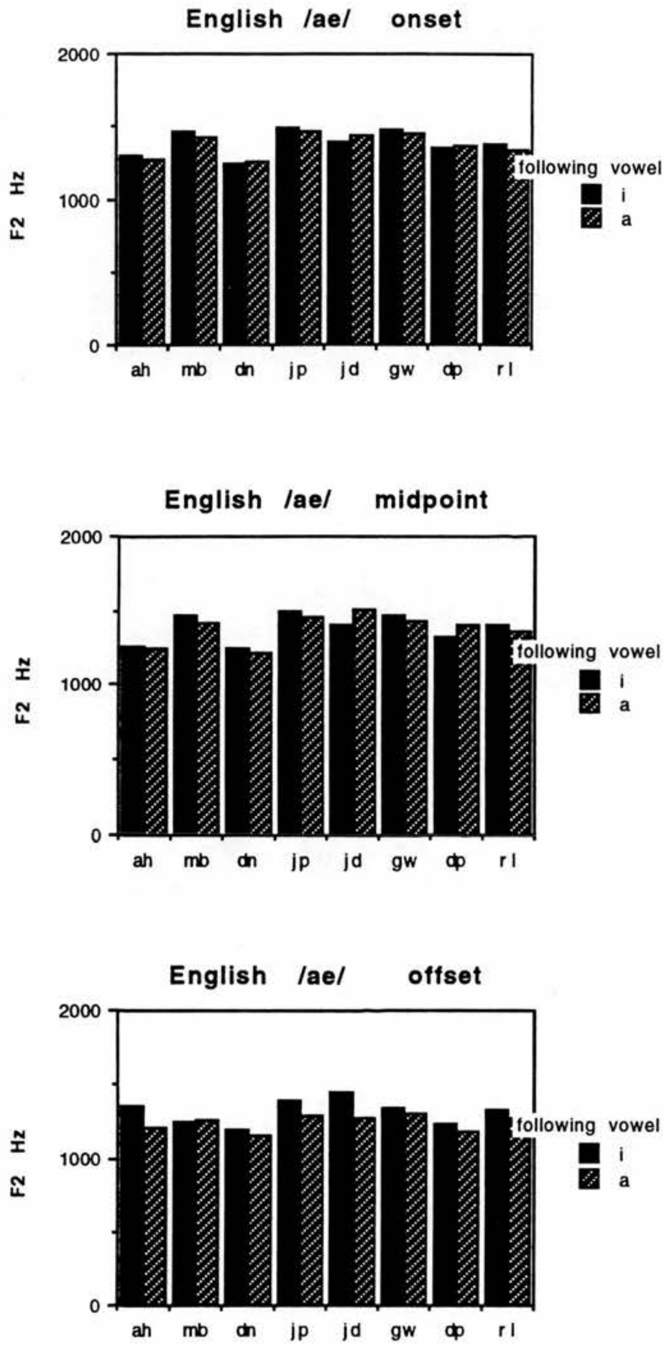
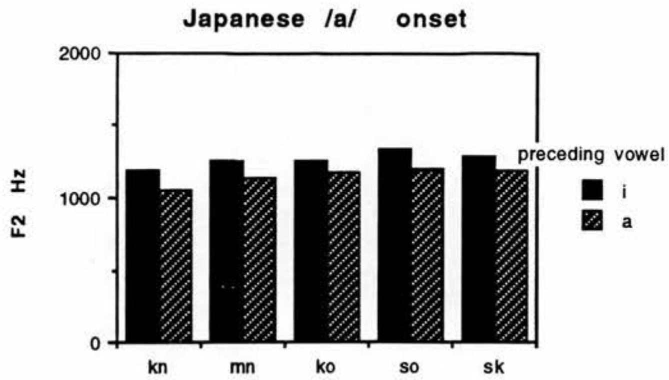
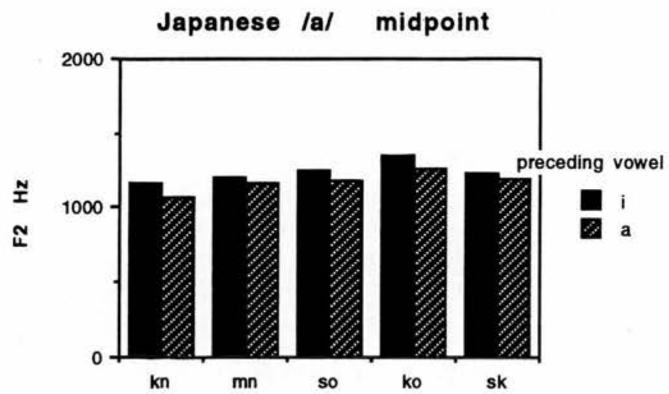

English /ae/ onset
F2 Hz
2000
1000
0
ah mb dn jp jd gw dp rl
following vowel
i
a
English /ae/ midpoint
F2 Hz
2000
1000
0
ah mb dn jp jd gw dp rl
following vowel
i
a
English /ae/ offset
F2 Hz
2000
1000
0
ah mb dn jp jd gw dp rl
following vowel
i
a

Figure A.4.


Japanese /a/ onset
2000
1000
0
F2 Hz
kn
mn
ko
so
sk
preceding vowel
i
a


Japanese /a/ midpoint
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
preceding vowel
i
a

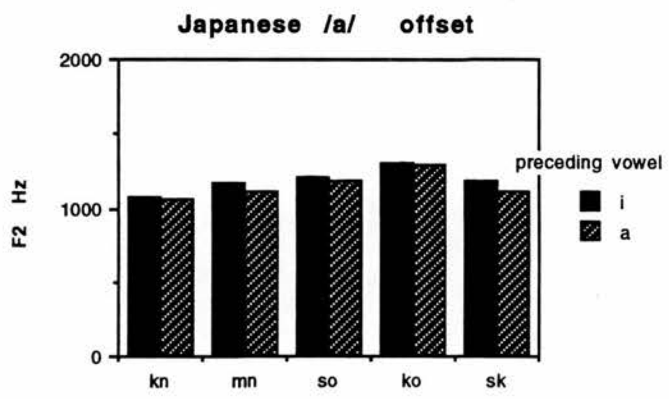

Japanese /a/ offset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
preceding vowel
i
a

Figure A.5.

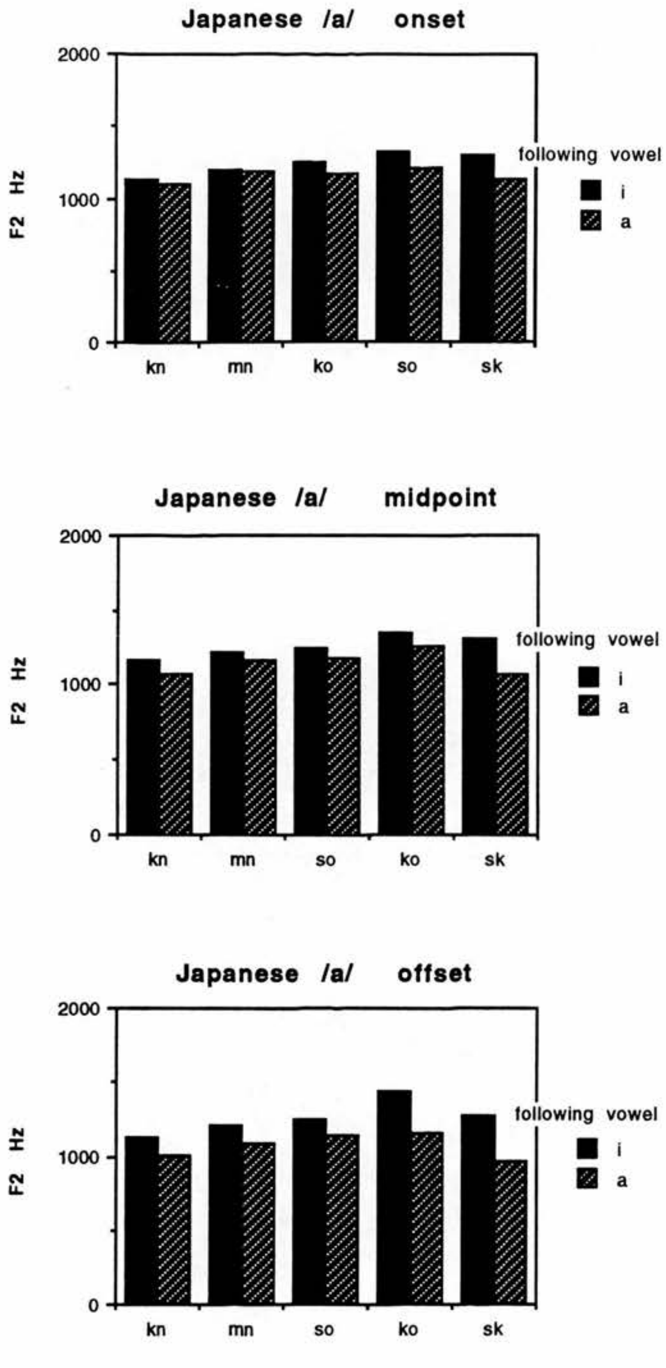
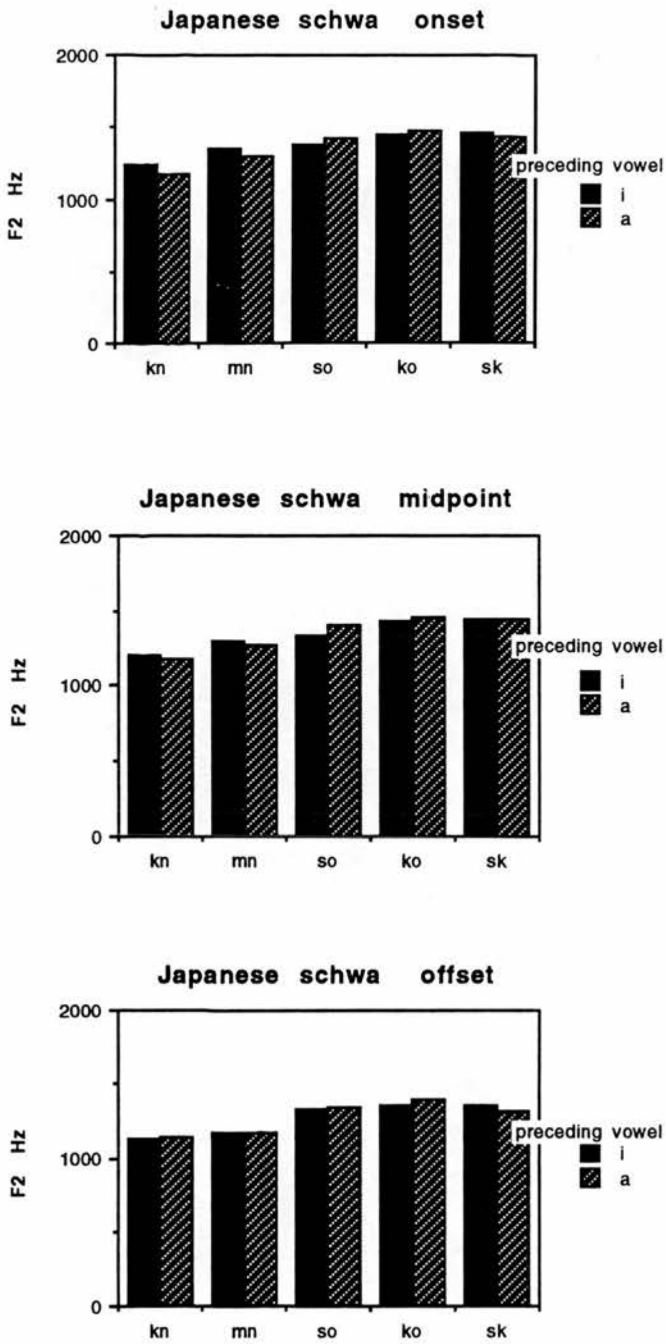

Japanese /a/ onset
2000
1000
0
F2 Hz
kn
mn
ko
so
sk
following vowel
i
a
Japanese /a/ midpoint
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a
Japanese /a/ offset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a

Figure A.6.



Japanese schwa onset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
preceding vowel
i
a
Japanese schwa midpoint
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
preceding vowel
i
a
Japanese schwa offset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
preceding vowel
i
a


**Figure A.7.**

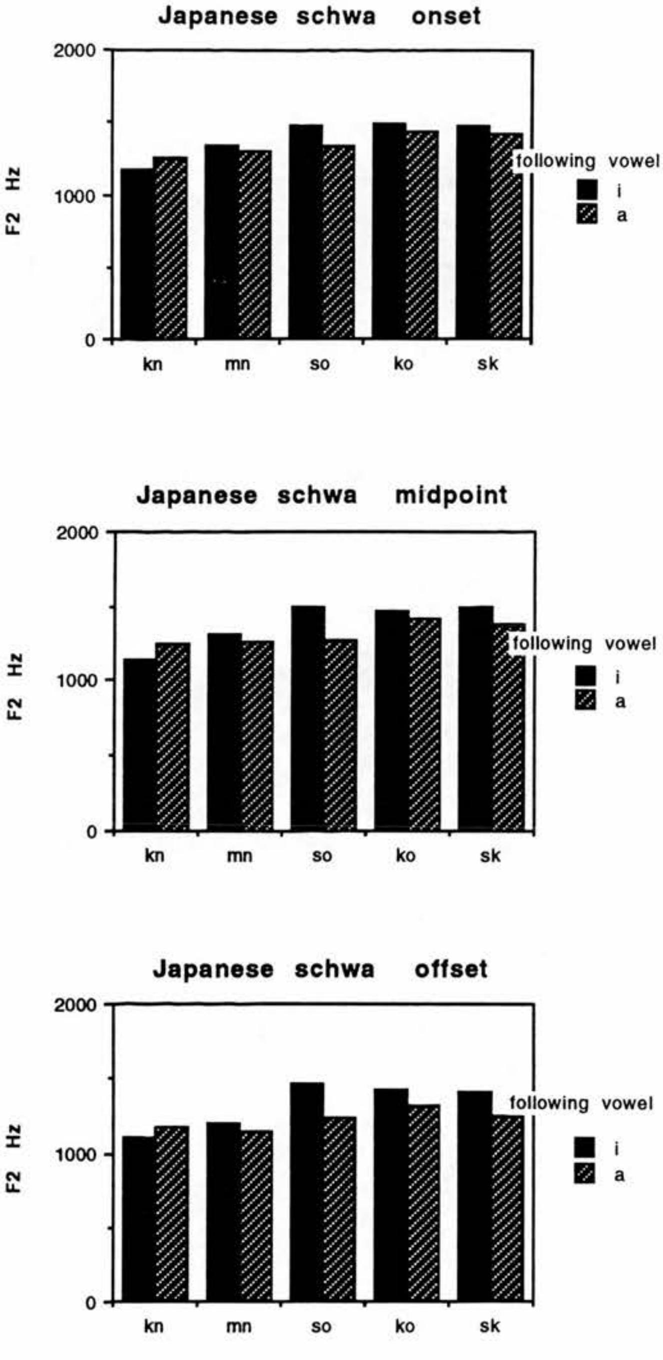

Japanese schwa onset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a
Japanese schwa midpoint
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a
Japanese schwa offset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a

**Figure A.8.**

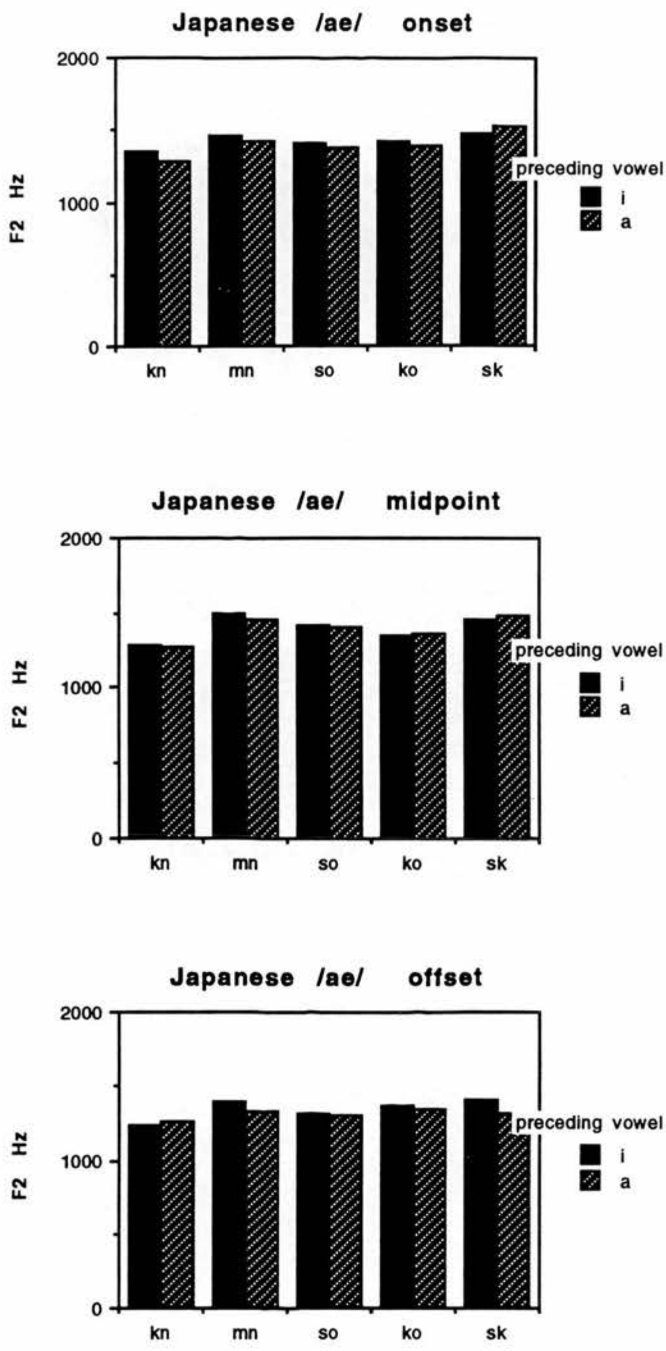

Japanese /ae/ onset
2000
1000
0
F2 Hz
kn mn so ko sk
preceding vowel
i
a
Japanese /ae/ midpoint
2000
1000
0
F2 Hz
kn mn so ko sk
preceding vowel
i
a
Japanese /ae/ offset
2000
1000
0
F2 Hz
kn mn so ko sk
preceding vowel
i
a

**Figure A.9.**

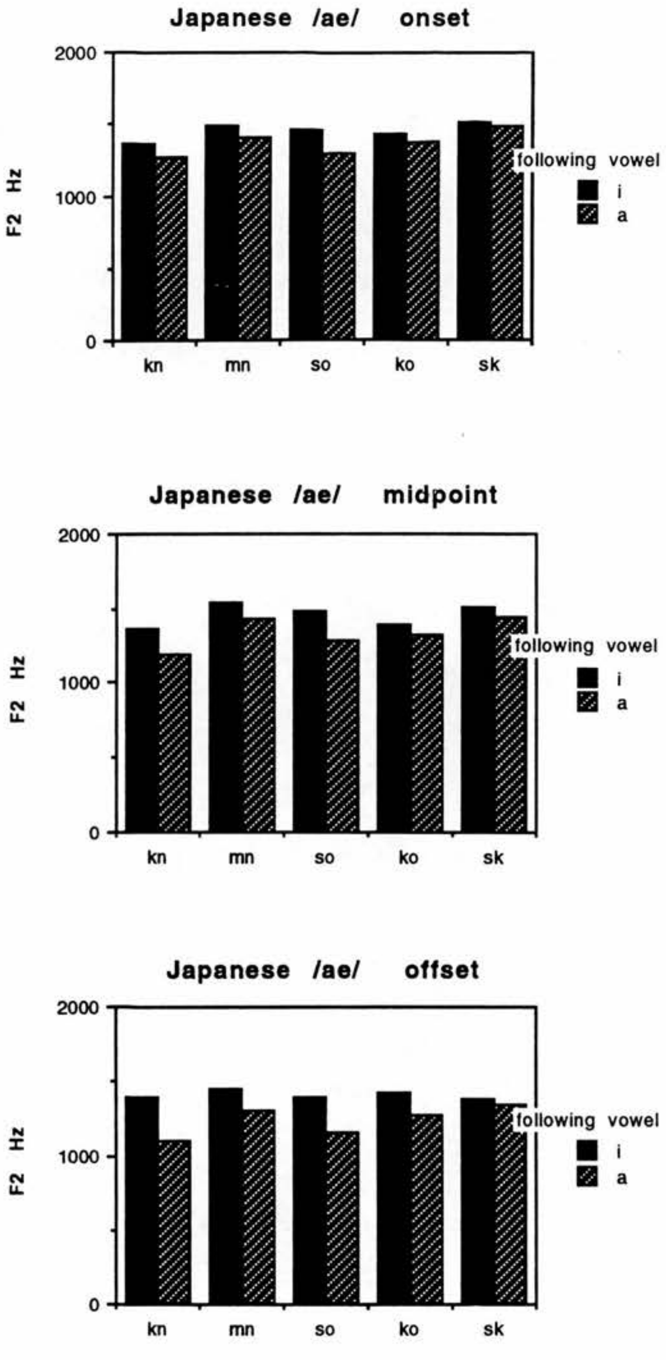

Japanese /ae/ onset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a
Japanese /ae/ midpoint
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a
Japanese /ae/ offset
2000
1000
0
F2 Hz
kn
mn
so
ko
sk
following vowel
i
a

Figure A.10.

# Appendix B

This appendix contains figures displaying the mean $F_2$ values in the four (symmetric and asymmetric) Vb_bV contexts, with the contextual vowels /ɪ/ (/i/) and /æ/ (/ə/ or /a/), at the onset, midpoint and offset of the English vowels /ə/ and /æ/ produced by native and non-native speakers of English and the Japanese vowel /a/ produced by Japanese subjects. The four contexts are /ɪbəbɪ/, /ɪbəbæ/, /æbəbɪ/ and /æbəbæ/ for schwa, /ɪbæbɪ/, /ɪbæbə/, /æbæbɪ/ and /æbæbə/ for the full vowel /æ/ and /ibabi/, /ibaba/, /ababi/ and /ababa/ for the Japanese vowel /a/.

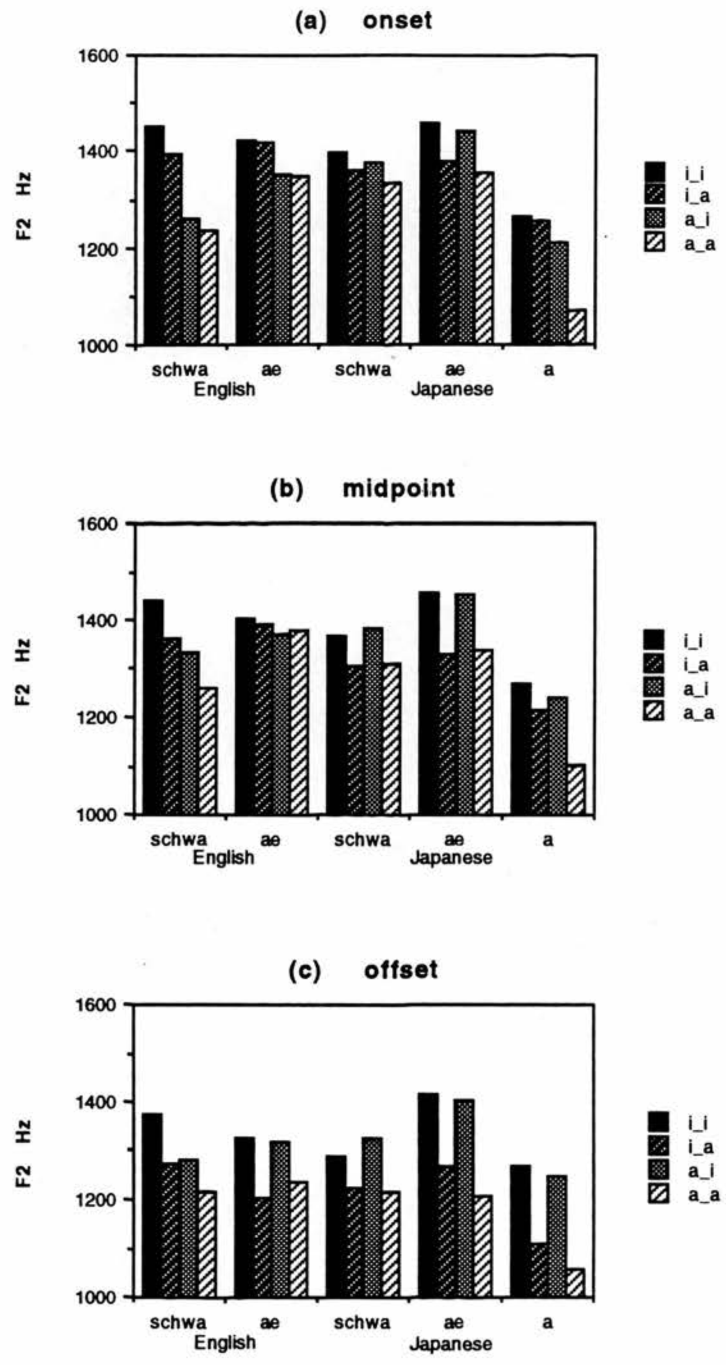

(a) onset
F2 Hz
1600
1400
1200
1000
schwa
ae
schwa
ae
a
English
Japanese
i_i
i_a
a_i
a_a
(b) midpoint
F2 Hz
1600
1400
1200
1000
schwa
ae
schwa
ae
a
English
Japanese
i_i
i_a
a_i
a_a
(c) offset
F2 Hz
1600
1400
1200
1000
schwa
ae
schwa
ae
a
English
Japanese
i_i
i_a
a_i
a_a

**Figure B.1.**

# Appendix C

This appendix displays a matrix of the $F_1$ and $F_2$ values as a function of the consonantal contexts /p, t, k/ and the vocalic contexts /ɪ, æ, u/ for the production of schwa by each Japanese speaker. The top three subjects, KN, MT and MN are fluent non-native speakers of English and the bottom three subjects, TT, HK and HS are non-fluent non-native speakers of English.

| $F_1$ | | | | | $F_2$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| KN | p | t | k | | KN | p | t | k | |
| ɪ | 328 | 354 | 329 | 338 | ɪ | 1195 | 1304 | 1740 | 1413 |
| æ | 294 | 394 | 315 | 338 | æ | 1150 | 1327 | 1629 | 1370 |
| u | 356 | 319 | 323 | 331 | u | 1168 | 1339 | 1281 | 1263 |
| | 326 | 354 | 322 | 336 | | 1171 | 1323 | 1550 | 1348 |
| MT | p | t | k | | MT | p | t | k | |
| ɪ | 375 | 389 | 319 | 361 | ɪ | 1553 | 1889 | 2164 | 1869 |
| æ | 343 | 475 | 319 | 379 | æ | 1419 | 1546 | 2134 | 1700 |
| u | 368 | 394 | 345 | 369 | u | 1233 | 1499 | 1666 | 1466 |
| | 362 | 422 | 328 | 370 | | 1401 | 1630 | 1991 | 1678 |
| MN | p | t | k | | MN | p | t | k | |
| ɪ | 483 | 431 | 434 | 449 | ɪ | 1349 | 1600 | 1669 | 1539 |
| æ | 494 | 470 | 427 | 463 | æ | 1350 | 1563 | 1651 | 1522 |
| u | 518 | 424 | 434 | 449 | u | 1233 | 1583 | 1470 | 1428 |
| | 495 | 441 | 432 | 454 | | 1311 | 1582 | 1597 | 1496 |
| TT | p | t | k | | TT | p | t | k | |
| ɪ | 276 | 325 | 503 | 368 | ɪ | 1345 | 1468 | 1489 | 1434 |
| æ | 304 | 433 | 322 | 353 | æ | 1217 | 1447 | 1288 | 1317 |
| u | 468 | 313 | 380 | 387 | u | 1253 | 1366 | 1317 | 1312 |
| | 349 | 357 | 403 | 369 | | 1272 | 1427 | 1365 | 1354 |
| HK | p | t | k | | HK | p | t | k | |
| ɪ | 542 | 387 | 447 | 438 | ɪ | 1310 | 1695 | 1602 | 1536 |
| æ | 484 | 339 | 400 | 398 | æ | 1289 | 1638 | 1445 | 1457 |
| u | 804 | 572 | 718 | 698 | u | 1334 | 1556 | 1532 | 1474 |
| | 655 | 436 | 522 | 524 | | 1311 | 1630 | 1526 | 1489 |
| HS | p | t | k | | HS | p | t | k | |
| ɪ | 279 | 272 | 260 | 270 | ɪ | 1302 | 1546 | 1412 | 1420 |
| æ | 283 | 393 | 290 | 333 | æ | 1302 | 1522 | 1409 | 1411 |
| u | 714 | 300 | 365 | 440 | u | 1303 | 1432 | 1346 | 1360 |
| | 425 | 321 | 299 | 343 | | 1302 | 1500 | 1389 | 1397 |

**Table C.1.** The mean $F_1$ and $F_2$ values of the vowel /ə/ measured at the vowel midpoint in the 3 consonant /p, t, k/ × 3 vowel /ɪ, æ, u/ contexts for each speaker. The rightmost column of each matrix shows the mean formant frequencies for the vocalic contexts. The bottom row shows the mean formant frequencies for the consonantal contexts. The rightmost value on the bottom row is the grand mean for each matrix.